DOCUMENT RESUME

ED 384 472 RC 020 175

TITLE	Assessment in Rural and Small Schools. Rural, Small Schools Network Information Exchange: Number 12, Spring 1992.
INSTITUTION	Regional Laboratory for Educational Improvement of the Northeast & Islands, Andover, MA.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	92
CONTRACT	RP91002008
NOTE	110p.; Photographs and some pages may not reproduce adequately.
PUB TYPE	Collected Works - General (020)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	Academic Achievement; Competency Based Education; Educational Change; Educational Practices; Educational Testing; Elementary Secondary Education; *Evaluation Methods; Informal Assessment; Portfolio Assessment; Principals; *Program Evaluation; *Rural Education; Self Evaluation (Groups); Self Evaluation (Individuals); *Student Evaluation; *Teacher Evaluation
IDENTIFIERS	Alternative Assessment; *Authentic Assessment

ABSTRACT

This packet includes reprints of journal articles and other information concerning student, program, and staff assessment in small, rural schools. The five sections of the packet cover an introduction to authentic assessment, classroom and subject area assessment, descriptions of assessment programs, assessment of school boards and staff members, and additional resources. Articles include: (1) "The Regional Lab Reports on Assessment"; (2) "Putting It in Perspective" (Richard Nicoll); (3) "The Use and Abuse of Test Scores: What Do the Results Really Tell Us?" (William J. Leary); (4) "Linking Restructuring to Authentic Student Achievement" (Fred M. Newmann); (5) "How Can a Student 'Fail' To Be Educated? (Phil Rogers); (6) Finding the Value in Evaluation: Self-Assessment in a Middle School Classroom" (Linda Rief); (7) "Show What You Know" (Deana R. Chadwell); (8) "On Assessment in the Arts: A Conversation with Howard Gardner" (Ron Brandt); (9) "Portfolio Assessment for Young Readers" (Sheila W. Valencia); (10) "Assessing Science Learning in Elementary School: Why, What, and How?" (Senta A. Raizen, Joyce S. Kaser); (11) " A Changing World Requires Changes in Math Instruction" (Gail Marshall); (12) "A Large-Scale Assessment To Support the Process Paradigm" (Charlotte W. O'Brien). (13) "Using Writing Portfolios: An Interview with John Collins" (Lori Mammen); (14) "An Alternative Assessment Program in Language Arts" (Dennis Searle, Margaret Stevenson); (15) "How To Create a Lifelong Love of Reading: Expert Shares Tips To Evaluate Your Reading Program" (Sidney J. Rauch); (16) "Can Cooperative Learning Be Evaluated? Yes, If the Observer Knows the Keys to Its Success" (Dan Watson, Lyle Rangel); (17) "Beginning Researchers: How To Make Report Writing More than Copying from World Book" (Donna Maxim); (18) "A School Board's Need for Self-Assessment" (Donald Overly, Darrell K. Root); (19) "Self-Assessment in the Evaluation Process" (Michael Koehler); (20) "How Best To Evaluate Principals? Spotlight on Site-Level Operations Sparks Keen Interest in Principal Evaluations" (Mark E. Anderson, Linda Lumsden); (21) "Educative Possibilities in Teacher Evaluation: Two Alternatives" (Andrew Gitlin, John Smyth); and (22) "Watching Teachers Teach" (Judith D. Clemente, Ruth B. Greenblatt, Robert E. Maher). Section 5 lists additional resources. (LP)

ED 384 472

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Janet Angelis

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

The Regional Laboratory
for Educational Improvement of the Northeast & Islands

Spring 1992

Dear Rural, Small School Leader:

This Information Exchange Packet on **Assessment in Rural and Small Schools** is the twelfth in a series of packets offered to members of the Rural, Small Schools Network. The articles chosen here are a sample of, and a supplement to, the many varied and rich resources on assessment available to educators today, looking at student, program, and staff assessment.

Information Exchange Packet #12, **Assessment in Rural and Small Schools** is divided into five sections. Section 1 provides an overview of concepts and issues surrounding authentic assessment. Section 2 looks at assessment at the classroom and subject area level. Section 3 takes a broader, programmatic view. Section 4 deals with assessment of school boards and staff members, and Section 5 provides a list of additional resources for those interested in further reading on the topic.

With The Regional Laboratory, some educators in the Northeast are working with assessment tools that address the intersection of assessment, teaching, and learning: Rural, Small Schools Network members will be kept informed about these resources as they become available.

An evaluation card has been provided to let us know how useful you have found this packet. We also welcome your suggestions for future Information Exchange Packet topics. Please jot down any ideas you may have on the card or contact us at the Rural, Small Schools Network, 83 Boston Post Road, Sudbury, MA 01776, (508) 443-7991.

Sincerely,

John R. Sullivan, Jr., Ed.D.
Program Director
Rural, Small Schools Network

Carolyn J. Burke, Ed.D.
Associate Program Director
Rural, Small Schools Network

RC020175

CONTENTS

ASSESSMENT IN RURAL AND SMALL SCHOOLS

SECTION 1:

"The Regional Lab Reports on assessment," Newsletter of The Regional Laboratory for Educational Improvement of the Northeast and Islands, January 1992.

"Putting it in PERSPECTIVE," by Richard Nicoll, Thrust for Educational Leadership, October 1991.

"The Use and Abuse of Test Scores, What do the results really tell us?," by William J. Leary, Ed.D., MASC Journal, Spring 1990.

"Linking Restructuring to Authentic Student Achievement," by Fred M. Newmann, Phi Delta Kappan, February 1991.

"How Can a Student 'Fail' To Be Educated?," by Phil Rogers, Phi Delta Kappan, February 1989.

SECTION 2:

"Finding the Value in Evaluation: Self-Assessment in a Middle School Classroom," by Linda Rief, Educational Leadership, March 1990.

"Show What You Know," by Deana R. Chadwell, The American School Board Journal, April 1991.

"On Assessment in the Arts: A Conversation with Howard Gardner," by Ron Brandt, Educational Leadership, December 1987/January 1988.

"Portfolio assessment for young readers," by Sheila W. Valencia, The Reading Teacher, Volume 44, No. 9, May 1991.

"Assessing Science Learning In Elementary School: Why, What, and How?," by Senta A. Raizen and Joyce S. Kaser, Phi Delta Kappan, May 1989.

"A changing world requires changes in math instruction," by Gail Marshall, The Executive Educator, July 1990.

"A Large-Scale Assessment to Support the Process Paradigm," by Charlotte W. O'Brien, English Journal, February 1992.

SECTION 3:

"Using Writing Portfolios: An Interview with John Collins," Writing Teacher, November 1991.

"An Alternative Assessment Program in Language Arts," by Dennis Searle and Margaret Stevenson, Language Arts, Volume 64, Number 3, March 1987.

"How to Create a Lifelong Love of Reading, Expert Shares Tips To Evaluate Your Reading Program," by Sidney J. Rauch, The School Administrator, May 1992.

"Can Cooperative Learning Be Evaluated? Yes, If the Observer Knows the Keys to Its Success," by Dan Watson and Lyle Rangel, The School Administrator, June 1989.

"BEGINNING RESEARCHERS, How To Make Report Writing More than Copying from World Book," by Donna Maxim, American Educator, Spring 1991.

SECTION 4:

"A School Board's Need for Self-Assessment," by Donald Overly and Darrell K. Root, MASC Journal, Fall 1990.

"Self-Assessment in the Evaluation Process," by Michael Koehler, NASSP Bulletin, September 1990.

"How Best To Evaluate Principals? Spotlight on Site-Level Operations Sparks Keen Interest in Principal Evaluations," by Mark E. Anderson and Linda Lumsden, The School Administrator, December 1989.

"Educative Possibilities in Teacher Evaluation: Two Alternatives," by Andrew Gitlin and John Smyth, NASSP Bulletin, September 1990.

"WATCHING TEACHERS TEACH," by Judith D. Clemente, Ruth B. Greenblatt, and Robert E. Maher, The Executive Educator, April 1992.

SECTION 5:

Additional Resources

SECTION 1

ASSESSMENT: AN OVERVIEW

The Regional Lab Reports

on assessment

Newsletter of The Regional Laboratory for Educational Improvement of the Northeast and Islands January 1992

Authentic Assessment:

The 'Trojan Horse' of School Reform

In Durham, New Hampshire, a middle-school student gazes at the score of "7" (out of 8) she has received on a district-mandated writing test. "What does this tell me?" she asks. "I'm one less than an 8, and one more than a 6. So what?"

[Linda Rief, Educational Leadership, *March 1990]*

In Gorham, Maine, a third-grade student's poem about autumn hangs on the wall. At the bottom she has written: "I think I deserved a [good mark] on this poem because I used figurative language, good penmanship, and it's not boring."

The contrast between these two students' reactions to their grades illustrates the close link between student learning and assessment: in one assessment the student apparently has gained no knowledge about what constitutes good writing; in the other, the student has learned a great deal. And what she has learned reflects what and how she has been taught. This example demonstrates the inseparable nature of teaching, learning, and assessment. Staff at The Regional Laboratory are exploring the intersection of these three fundamentals of education and the implications for developing *authentic* assesssment tools.

Defining Student Success

The first step in any assessment is to define student success, a definition that depends to a large extent on knowing how students learn, what the collective school community wants students to know and be able to do, and what successful learning strategies look like in action.

Traditionally, however, thinking on assessment has driven thinking on teaching, which has driven thinking on student learning, in a linear type of relationship. In some schools, this linear relationship continues to encourage the attitude that "testing is legitimated by someone outside the student, teacher, school, and community," says Lab staff member John Watkins. "We then separate the curriculum in ways that seem logical to textbook publishers or test publishers, whose decisions some educators increasingly come to rely on because we no longer feel we can be experts on anything we teach," building a greater reliance on standardized tests.

> One day assessment may even come to be viewed as what it is—a tool of learning, not a weapon of control.

Standardized tests have been taken to task for many other shortcomings, including encouraging teachers to waste time (and even "dumb down" instruction) by "teaching to the test." But perhaps their biggest drawbacks are that they measure only limited skills, such as short-term recall, instead of real knowledge or thinking skills, and that they fail to tell schools anything useful about student learning.

"There is no feedback," says Watkins, "from learning to designing assignments, from teaching to assessment, from how kids learn to how teaching happens—not to mention

(continued on page 4)

National Testing System Proposal Draws Fire

It's not surprising that many educators are alarmed that at the very moment when standardized testing at the local level is coming under fire, the Bush administration and others are urging more national testing.

The National Education Goals Panel (NEGP), established to measure the nation's progress toward the National Education Goals adopted by President Bush and the nation's governors in 1989, relied on a variety of measures in issuing its first "report card" this September. However, the panel was unable to find a single set of data to measure progress toward any one of the goals. As a result, it has called for the creation of national standards and a new national assessment system to determine how many students meet the new standards. The NEGP is comprised of governors, members of the Bush administration, and members of Congress (ex officio).

The Panel also sought the establishment of a National Council on Education Standards and Testing (NCEST), which last month recommended expanding the NEGP to include voting members of Congress and more governors, with fewer administration officials. This new NEGP, with the aid of a technical council, would also oversee the development of voluntary national standards and assessments—to be created by individual states or in regional clusters.

The vision of regionally developed assessments tied to national standards is espoused primarily by Lauren Resnick, Director of the Learning Research and Development Center at the University of Pitts-

(continued on page 2)

National Testing

(continued from page 1)

burgh, and Marc Tucker, Director of the National Center on Education and the Economy.

Part of the administration's America 2000 strategy, unveiled in April 1991, calls for the adoption of national standards and the creation of American Achievement Tests. These tests would be tied to the National Education Goals and would be voluntary in nature. As proposed, they would serve many disparate functions simultaneously: to hold schools accountable; to foster good teaching and good learning; and to provide a basis for hiring and college admissions.

Educators have been fairly critical in their response to the idea, saying that one achievement test cannot be used to meet the three different stated goals and that using the tests to hold schools accountable, with a system of rewards and sanctions, would have ill effects for schools that serve the most at-risk students.

Though some who seek to establish a national testing system call authentic assessment methods to be a part of that, other educators worry that such a test will encourage even more the practice of "teaching to the test," which will narrow the curriculum and may even prescribe a de facto national curriculum. Other fallout from the tests could be a set of standards that, instead of being "world-class," as the Panel hopes, meet only the needs of low performing states.

Guest Editorial

Beware the Bell-Shaped Curve

by Raymond Rose
Director, the New England Center for Equity Assistance

Do you remember that class in which your teacher "graded on a curve?" Even though you got most of the answers right, your classmates who did better got the As and Bs. Or maybe you've been saved by the curve: although you may have done poorly on a test, you out-performed your classmates, so you got the A or B.

Most educators are familiar with the normal bell-shaped curve. But how often do we stop to ask where it came from, how we use it, if it's valid, or even if it's "normal?"

When Alfred Binet planned his I.Q. test some 80 years ago, he predetermined that the distribution of intelligence among people should fall along a bell-shaped curve. And so he constructed a test that would produce such results. In fact, in order to be considered valid, he felt, his test had to produce the curve.

Many test makers today are still doing so. Test publishers develop tests by first creating a large number of test questions, which they administer to a statistically large population. Based on the results, they then select that combination of questions that will wield a predetermined distribution, most often the normal bell-shaped curve.

What's the problem? you might ask. I see two. First, we have never determined that Binet's assumption about the distribution of intelligence among the population is valid. It seems to me that Binet created a chicken and egg riddle: If the instruments we use to measure intelligence are based on assumptions about the distribution of intelligence (because there is no measure), then how do we know that the results we get from such tests are really measuring intelligence?

The second problem is usage. We damage the psyche of too many students when test results that were *set up* to produce the bell-shaped curve are used to make absolute judgements about their abilities. Our judgement—perceived by students in the way we treat and teach them—convinces them that they are not capable.

When I was a guidance counselor there were a number of teachers whose first action when assigned a new class was to check student test records. Often, a teacher would approach me to transfer a particular student to another class because he or she had determined—by looking at the test scores—that this student was incapable of performing well in that class, in spite of evidence that the student had held his or her own among that group of students in the past. If the teacher was not granted the transfer, invariably the student did poorly. I was always left wondering if the cause was the ability of the student or the teacher's expectations.

One reason that I care so much about this problem is that I see too often which students are generally at the bottom—minority kids. A number of years ago a group of black psychologists produced an I.Q. test called the "Chitlin' Test" to illustrate the cultural bias inherent in any test. They developed and normed the test based on the African American cultural experience. As you might expect, that portion of the population that does not share this cultural experience fell to the bottom of the curve.

Another example in which a test developer has deliberately built bias is the Scholastic Aptitude Test (SAT). Historically males have outscored females on the math portion of that test by about forty points. Questions are deliberately chosen so that the test continues to produce this disparity. Based on test trials, questions that produce higher female scores are discarded. Thus the developers ensure that the SAT will produce the result they expect.

If we are really serious about education reform, we must look not only at the basic assumption of tests (e.g., that males will outperform females in math, that intelligence is distributed along a bell-shaped curve) but at how we use the results. If schools are going to accept test results as the true and only measure of a child's intelligence, then I say get rid of the tests. Tests that grade on a bell-shaped curve should never be the sole or primary basis on which to make decisions about individual students. Rather, let's look at the real potential of each child and provide challenging educational opportunities for all.

The Executive Director's column will return to this space with the next edition of **The Regional Lab Reports.**

Changes in Testing Around the Northeast: A Sampling of Authentic Assessment

Many different authentic tests have already been designed and are being tried. In our region, a number of schools, projects, and even an entire state, are leading the way, among them:

• **The Portfolio Project.** The nation's first state-level study of student assessment using 4th and 8th grade portfolios is currently taking place in Vermont. Over 135 schools, including 40 taking part in a pilot program, converted to a portfolio system of assessment last year for writing and mathematics. The results, released in July 1991, show that "the portfolio system worked," according to Education Commissioner Richard P. Mills. Not only were standards "demanding," he said, but the portfolio system helped educators pinpoint areas of weakness, especially in mathematics. The Vermont Department of Education plans to train teachers in new math strategies and hopes to have portfolio testing fully implemented in the two grades statewide by 1992.

Networking. The Vermont Department of Education has created networks through which fourth- and eighth-grade teachers and administrators can become informed about opportunities to learn more about portfolio assessment and how to develop such a system; in addition, multiple *computer* networks allow them to exchange ideas. "Teachers are laying the foundation for [statewide idea-swapping] right now," says the DOE, by using computer bulletin board notices, e.g., "Does anyone have good problem sets I can use to illustrate this concept...?"

• **Performance Tasks and Group Interviews.** In 1989 selected schools in the districts of East Providence, Glocester, South Kingstown, and Newport, Rhode Island began a pilot assessment project with the state department of education and Educational Testing Service (ETS). The inital goal of the project was to create a set of standard outcomes for the state's third graders through the use of portfolios. Since the project's inception, Rhode Island, educators have incorporated other methods of evaluation such as *performance task assessment*—traditionally used to evaluate students' abilities in performance and multi-media arts and sports; and *group interview assessment*—a method which allows teachers to facilitate and observe students as they take part in topical discussions (*Phi Delta Kappan*, December 1991).

• **Portfolios for Large-Scale Use.** Oyster River High School in Durham, New Hampshire, has successfully pilot-tested an alternative to timed writing samples. Using a portfolio method, that measures production, perceptions, and reflection in fifth grade writers, they compared their results with the same students' timed test scores. "Our test results indicate that self-selected portfolios are significantly better than timed tests in estimating student writing abilities," concluded team member and English teacher Jay Simmons, "especially with respect to weaker writers (*Educational Leadership*, March 1990)."

> "Our test results indicate that self selected portfolios are significantly better than timed tests in estimating student writing abilities, especially with respect to weaker writers."
>
> *Jay Simmons, English Teacher, Oyster River High School, Durahm, NH*

• **Curriculum Building and Assessment "Mirroring."** The Narragansett School (K-3) in Gorham, Maine, is working on curriculum building and assessment "mirroring" as part of an ongoing restructuring effort. Students use portfolios, which they annotate, an "extremely important" part of the portfolio idea, according to principal Cindy O'Shea. The school is also emphasizing the use of scoring guides (a set of standards that will allow students to see how and why their work is scored). The school has applied for a grant to work with the local high school on developing the scoring guides. Perhaps the most innovative move is their attempt to capture portfolios using electronic scanners; so far they have created several such prototypes.

• **Program Assessment.** The New York State Department of Education is conducting a large-scale statewide assessment of K-4 science programs with seven instruments, of which two are required—a written test and a performance test. Other instruments include surveys of student attitudes, the classroom environment, and teachers, principals, parents, and teachers. The assessment effort, in which 200,000 students have participated, is designed to measure science learning, tie assessment to the state syllabus, and provide each school with an analysis that identifies which strategies are working well.

(continued from page 1)

how kids judge themselves."

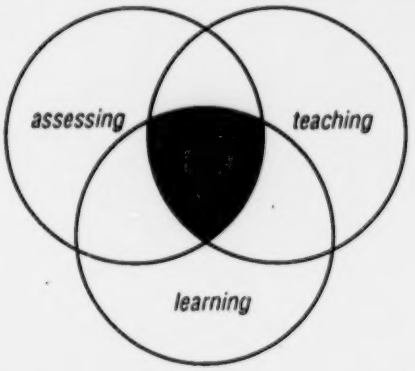
Connecting Assessment with Instruction

Many educators are so busy trying to keep up with the external demands of testing, they often miss the integral connection there should be to teaching and learning: "We haven't had any time to work on thinking, we've been so busy testing," one teacher recently told the Lab.

The traditional linear view of assessment should be replaced by looking at testing, teaching, and learning as "overlapping sets" and examining the areas of intersection. Testing will then no longer be an outside force, but an integral part of a school. Lab staffer Pat Cox warns that "you can't just graft on a new way of testing—nor can you keep the old way [of teaching] when you have shifted ways of assessing."

A number of broad questions must be tackled before designing new assessments, for instance: What is learning? What is success? And what does it mean for all learners to learn successfully?

"Assessment is the Trojan horse of school reform," says Grant Wiggins, because it drives instruction. Wiggins, a consultant on assessment for the Coalition of Essential Schools and the National Center on Education and Economy, says that schools must first do some soul searching about what they want students to learn, and then design tests: "Do we judge our students to be deficient in writing, speaking, listening, artistic creation, research, thoughtful analysis, problem-posing, and problem-solving? "Let the tests ask them to write, speak, listen, create, do original research, analyze, pose and solve problems (*Educational Leadership*, April 1989)."



Authentic assessment takes place at the intersect of learning, teaching, and assessing.

This, he argues, will do away with the problem of "teaching to the test," because the test will become a legitimate goal: "Rather than seeing tests as only after-the-fact devices for checking up on what students have learned, tests should be seen as instructional, the central vehicle for clarifying and setting intellectual standards. The recital, debate, play, or game. . . the 'performance' is not a check-up, it is the heart of the matter, all coaches happily teach to it."

For example, concurs Cox, a third grader learning to write makes "books" during the school year. The books are both the "learning tool" and the assessment tool. The general principle, says Cox, is that "what is done in the classroom can be part of the assessment."

Designing Authentic Assessment

Authentic tests should provide real indicators of students' progress toward desired outcomes. Wiggi[illegible] says authentic tests have four bas[illegible] characteristics:

(1) *Structure.* Authentic tests are designed to be "truly representative of performance in the field."

(2) *Intellectual design features.* Authentic tests emphasize learning the

(continued on page 5)

Types of Traditional Tests

To help non-educators sort through the wide variety of tests used in K-12 and post secondary education, the National PTA created a summary chart of common assessment tools, adapted below for The Regional Lab Reports:

- *Standardized tests* are identical tests designed to be given in thousands of classrooms at different times under standard conditions. Some examples are the SATs and the Iowa Tests of Basic Skills. They allow the comparison of students in schools throughout the country.
- *Achievement tests* are used to measure what a child has learned in a subject. Final exams in the classroom are kinds of achievement tests.
- *Classroom (or teacher-created) tests* are used to find out how much students have learned of what the teacher has taught, to identify strengths or weaknesses, motivate students, measure progress, and award grades.
- *Aptitude tests* measure skills and abilities that are less closely linked to specific subjects but measure capabilities useful in learning new things. They are used to predict how a student will do in school, identify gifted children, or help in career planning.
- *Norm-referenced tests* are graded on a "curve" to allow each student to be measured against the performance of other students across the state, region, or country. The California Achievement Test, for example, shows what percentile each student falls into.
- *Criterion-referenced tests,* on the other hand, focus on standards, criteria, and benchmarks. For example, such a test would show how well a child has learned to multiply fractions, regardless of how well other children do it. A criterion-referenced test might tell you that a student met 11 out of 15 instructional objectives in math, while a norm-referenced test might tell you that this child pe[illegible] formed better than 75% of the ch[illegible] dren in the state.

(*Source: Adapted from* "Plain Talk about Tests," *National PTA, 700 North Rush Street, Chicago, IL 60611-2751*)

(continued from page 4)

criteria used in the assessment, enabling the student toward a better use of knowledge.

(3) *Standards of scoring.* Authentic tests measure essentials, not easy-to-count errors; they give a far greater role to self-assessment than usual; they do not incorporate grading curves; and their scoring systems are complex and multi-faceted.

(4) *Fairness.* Authentic tests often require a public presentation or defense to assure genuine mastery, allowing each student to show off what he or she can do best.

To be authentic, assessment must be of use in the teaching and learning processes, and conducted in multiple ways. Types of authentic assessment include portfolios, exhibitions of mastery, open-ended tests, and various kinds of projects or performances.

Knowing the Criteria in Advance

One example Wiggins offers is an oral history project for ninth graders at Hope High School in Providence, Rhode Island. Students are given sample topics and asked to prepare an oral history based on interviews and written research, which is then presented in class. They are told to create three workable hypotheses from their investigations and four questions to test the hypotheses.

Hope High School students receive specific criteria the teacher will use to assess their work. The list might read: Were three hypotheses investigated? Was at least one change over time described? Was there a clear demonstration of background research? Were four appropriate sources interviewed? Were unbiased questions and follow-up questions asked? Were the differences between fact and opinion noted? Was evidence used to support their choice of hypothesis? And, was the presentation organized? (*Educational Leadership*, April 1989)

By knowing criteria in advance, students are able to judge themselves as they go along.

Self-assessment is clearly an important part of any authentic test. Many portfolio approaches require students to select their own best work samples and to evaluate their weak areas.

Obstacles

In part, standardized testing came about in the early part of the century because of a perception that the school transcript was not reliable. An "A" in English from one school conveyed nothing to another school, college, or employer. Today, as educators move towards more non-standardized assessment they are faced with the challenge that reliability of scoring and authenticity of material continue to be seen as a problem by others.

There are plenty of other reasons why authentic assessment will be difficult to implement. The most obvious is cost, both in dollars and teacher time. Computerized scoring is far cheaper than systems that require teachers to personally evaluate students. According to Bruce C. Bowers, "labor intensive" scoring systems can be "prohibitively expensive" and must justify the cost in terms of educational gains (*ERIC Digest*, EA 40 1989).

And, opponents argue, since the concepts are still new, there has been relatively little research and development on assessment tools.

Wiggins refutes these claims, asserting that "adequately reliable ways of assessing knowledge have existed for decades and are used on a wide scale in other countries." In fact, he points out the United States is the only major country in the world to rely so heavily on norm-referenced tests. "The issue is not whether the 'testing technology' exists to permit fair grading of rich, authentic tests, but whether we will invest the time and money it takes to administer them (*RADIUS*, American Federation of Teachers, Winter 1988)."

Tools, Not Weapons

Authentic assessment measures will return a degree of control of the curriculum to the school, the teacher, and the community, and in doing so should diminish the frightening and often baffling process of test taking for students. Authentic assessment will result in meaningful, useful evidence of what and how students learn and what and how teachers teach. And one day assessment may even come to be viewed as what it is—a tool of learning, not a weapon of control.

Portfolios should function as . . .

A collection of students' best work ***not as*** a collection of each week's papers.

A means for interaction between teacher and student ***not as*** the result of random, subjective choices.

A reflection of ongoing students' activities ***not as*** short-answer responses to textual questions.

A vehicle to involve students in reflecting about their own work ***not as*** evidence of teacher's evaluation only.

A basis for teacher/student conferencing ***not as*** the only means for teacher evaluation.

A vehicle for parent involvement ***not as*** a vehicle for infrequent parent conferences.

A shared teacher/student responsibility for organization and upkeep ***not as*** the teacher's responsibility for maintenance and upkeep.

A changing instrument to reflect growth ***not as*** a rigid design with set standards to achieve a product.

One form of assessment in conjunction with others ***not as*** the only form of assessment.

(Source: Draft document created by Alberta Patch-Slegaitis, Bureau Chief, Department of English Language Arts, NYS Education Department)

Lab Highlights

More than 60 educators came together on November 4 to honor the 1991-1992 Regional Laboratory Fellows at the Fourth Annual Teacher Recognition Forum. Teachers from rural and small schools were recognized for their exemplary work through a program conducted by the Lab's Rural, Small Schools Network. The event, held in Saratoga Springs, NY, honored Fellows and Finalists for their outstanding work in multilevel grouping in grades preschool through five.

Each fellowship includes a $500 honorarium for professional development of the honoree's choice. In addition, Fellows' exemplary programs are excerpted in the Lab's *Outstanding Teaching Practices* series.

In 1992-93, we will honor teachers in grades 6-12 for their exemplary work using research to improve education. For more information on the nomination process, contact Carolyn Burke at the Rural, Small Schools Network, (508) 443-7991.

Above are this year's Fellows, including (from top to bottom rows and left to right): Linda Deschenes, Lauren Potter, Susan Stellar (all from Lovell, ME), Barbara Holland Moody (Marion, MA), Karen Johnson (Lovell, ME), Karen H. Eaton (Geneseo, NY), Dennis Bathory Kitz (Roxbury, VT), Ellen Thompson (Colchester, VT), and Rhonda Poliquin (Lovell, ME).

The Regional Lab Reports
January 1992
A publication of
The Regional Laboratory
for Educational Improvement
of the Northeast and Islands

David P. Crandall, Executive Director
Janet Angelis, Editor
Jill Kaufman, Associate Editor
Amy MacDonald, Contributing Editor
Contributors to this issue: Pat Cox, Amy MacDonald, Raymond Rose, John Watkins

Bringing the best of educational research and practice to educators in Connecticut, Maine, Massachusetts, New Hampshire, New York, Puerto Rico, Rhode Island, Vermont, and the Virgin Islands.

300 Brickstone Square, Suite 900
Andover, Massachusetts 01810
Tel. (508) 470-0098
Fax (508) 475-9220

This publication is sponsored wholly or in part by the U.S. Department of Education, Office of Educational Research and Improvement, under contract number RP-91-002-008. The content of this publication does not necessarily reflect the views of the department or any other agency of the U.S. government.

The Regional Laboratory is an affirmative action employer.

Inside

- The "Trojan Horse" of Reform
- Beware the Bell-Shaped Curve
- National Testing
- Assessment Around the Northeast

Nonprofit organization postage **PAID** Permit No 69 Andover, MA

Route to: _____ _____ _____

Reprinting of the material contained in **The Regional Lab Reports** is encouraged. Please credit The Regional Laboratory for Educational Improvement of the Northeast and Islands.

Reprinted with permission from Thrust for Educational Leadership, October 1991.

Putting it in PERSPECTIVE

While touted as the answer to many of the dilemmas posed by student testing, performance-based assessment has its own shortcomings. Shades of gray in the standardized vs. authentic testing debate make a clear-cut choice between the two nearly impossible.

I believe in performance assessment. Last year my district tested the reading comprehension and math problem-solving abilities of all third and sixth grade students using performance assessment strategies. We have spent three years developing alternative measures, and we have a strong district commitment to them.

As I read the professional literature and listen to my colleagues, however, I am concerned about the one-sided nature of much of the present discussion about assessment. Current thought suggests there is little right with traditional, norm-referenced testing, and little wrong with alternative assessment. While this black and white distinction may help advance reform, a middle ground may prove more helpful in the long run. Both traditional and alternative models can contribute to improving the measurement of student achievement.

What would happen if the assessment tools available to physicians were judged with the same standards as norm-referenced tests? Clearly, the thermometer, stethoscope and blood pressure gauge are inadequate measures of a person's overall health. These three instruments may fail to detect a serious, life-threatening illness. Medical diagnosis is much better when the assessment is more complex. When assessments are recorded over time in a patient's folder (portfolio), the quality of diagnosis is further improved. Yet we don't hear about

by Richard Nicoll

thermometer bashing. Physicians accept crude instruments for what they are and conduct more thorough observations as needed.

The end of standardized tests?

Although test bashing is not new, it has never been so widespread. Some believe that the new assessments will eliminate many of the problems of "standardized" tests. What this argument misses is that performance assessments can be and already are "standardized," including most large-scale writing assessments such as the CAP direct writing exams, the New York State Elementary Science Program Evaluation Test and the Connecticut Assessment of Educational Progress.

Standardized simply means that each student has an equal opportunity to demonstrate understanding of the subject matter. For administrators to make comparisons between students, schools and/or districts as the current generation of educational reformers requires, students must have fair test administration and teachers must be consistent in how they judge student work. Standardization refers to how an assessment task is administered more than the nature of the task itself.

When I hear calls for performance assessment to replace all current norm-referenced testing, I have some doubts. We must proceed cautiously in applying costly assessment systems during a time of massive budget reductions in direct services to students. I am not ready to discard current testing instruments until I am confident that the new assessment tools work.

New assessment techniques must be able to stand up to the same level of criticism given to traditional tests because at some point the honeymoon with performance assessment will end. I am not suggesting that we reduce our efforts to improve student assessment, but before we bury our current testing tools we should make sure that we have something better in hand.

The road to Lake Wobegon, where all the children are above average, doesn't end with new assessments.

Performance assessment

Performance assessment has the potential to contribute greatly to educational measurement. In the long run, this type of assessment probably will have a positive impact on the classroom. The idea that "What you test is what you get" has been well accepted. There is clear evidence of this principle in our current testing system. Look no further than the prosperous test preparation materials industry to find evidence of testing affecting instruction. Some teachers now spend considerable classroom time drilling students on how to succeed on bubble tests.

I'm not so sure, however, that performance assessment by itself can improve student learning. Teaching to the current method of testing can have a negative influence in the classroom, but there is little evidence to suggest that improved assessment alone will improve instruction. Change in instruction also requires, at a minimum, staff development, support for teachers and principals who take risks, and time.

One study in England (Torrance 1991) tested the thesis that measurement can drive instruction. Even in a country regarded as highly progressive in its application of alternative assessment, the author found "teachers treated [performance] assessment as a special activity, set apart from teaching." The author concluded that after the different type of assessment, many teachers returned to teaching students with the same methods and materials they had used previously.

Because of its high visibility, reform of assessment can be thought of as a quick solution to the perceived need "to do something" to improve education. Let's not overlook the other components of educational change.

A related notion is that changing assessment systems will raise the scores of groups of students who have traditionally fared poorly on norm-referenced tests. This argument assumes that if the new assessment is relevant and meaningful, students will be able to demonstrate the full extent of their understanding. Because the new assessments are not norm-referenced, half of the students will no longer have to be below average.

In my district, we have found that students who score high on performance assessments also score high on norm-referenced tests. This is not a perfect relationship, but it is a consistent one. The achievement gap between the educationally privileged and the deprived has been well documented and probably will not disappear because of new tests. The road to Lake Wobegon, where all the children are above average, doesn't end with new assessments.

Some results of performance assessments indicate that students are achieving at lower levels than expected. In a recent statewide portfolio project, only one-fourth of the fourth grade students in Vermont showed evidence of reasoned decision-making in mathematics. This finding, of course, speaks to the value of performance assessment. Vermont educators are now aware of shortcomings in student achievement which they probably would never have identified using traditional testing strategies.

Finding a balance

Performance assessment will undoubtedly make enduring changes in the measurement of student achievement. For example, direct writing assessment is now widely accepted as the preferred

method of determining how well students can express themselves on paper. Yet 10 years ago, direct writing assessment was a radical innovation. I look forward to the time when direct assessments of math, science, and other academic disciplines are thought of as routine and practical.

Performance assessment will make a lasting contribution in its ability to assess directly the most important type of student work. We will no longer have to rely on a proxy for thinking skills. We will be able to model the type of classroom experiences we believe are most valuable for students. "Testing" will be less threatening, more integrated with everyday classroom work, and at times invisible to students.

I can see a time, however, when teachers may ask whether we need to spend so much time assessing students. Budget managers may wonder how to pay for the teacher time necessary to score performance assessments. At some point, someone may ask whether all student work must be judged this way, or whether there are some academic areas which can be assessed differently. The current emphasis on teaching thinking skills is not intended to mean that subordinate skills such as grammar and computation should not also be taught.

Some subjects can be properly assessed by traditional tests. Using a mixture of assessment techniques would permit the time available for performance assessments to focus on the most critical instructional tasks. As an example, the new CAP book, *A Sampler of Mathematics Assessment*, describes the use of "enhanced multiple choice items" among its four assessment options (California Department of Education 1991). CAP recognizes that multiple choice items need not be limited to trivial tasks and realizes their use in statewide assessments.

Using a variety of assessment techniques would allow teachers to focus on the most critical tasks in the the time available for performance assessments.

Computerized testing

Desktop computers could be invaluable in improving this type of assessment. For years, administrators of individual tests have selected appropriate items for each student based on the student's performance on other test items. Students who pass items get harder items; students who fail items get easier items. The examiner typically looks for the testing "floor," at which point the student gets everything right, and then locates the testing "ceiling," where the student can no longer answer questions. Individual examinations done in this manner do not require that all students take every item on the test. Norm-referenced tests, however, do. Top students and weak students take the exact same test, producing frustration for the weak students and boredom for the top students.

Computer adaptive testing is able to adjust the test based on student responses. Just as with individual tests, computer adaptive testing uses knowledge of previous student answers to select additional items. Each student attempts only items which are appropriate to the current level of performance. Testing time is reduced; there is no need for bubbling, scanning or any paper products; and students may take the test at any time. Teachers instantly know results. Sound too good to be true? The Northwest Educational Association already has a computer adaptive testing system on the market. As the cost of hardware decreases and more software becomes available, this type of assessment will increase.

The desktop computer can also help with performance assessment in other academic areas. Think of the potential risk of handing typical high school seniors five chemicals and asking them to conduct an experiment. The five chemicals form 10 unique pairs. Assume that nine of these combinations are harmless, but the tenth could, in the right circumstances, be dangerous.

Truly, open-ended student-directed performance tasks may be difficult in some academic disciplines. Instead of giving a student the chemicals, put the student at a computer with a chemistry program which allows them to replicate the experiment on the screen. The advantage, of course, is that the computer will not explode if the

wrong chemicals are mixed. This strategy is called computer simulated assessment.

Computer simulated assessment offers the potential to bring the power of integrated media to the assessment process. Imagine a task to assess knowledge of geology in which the students actually see and hear an earthquake of their own creation.

Conclusions

This is an exciting time to be working in student assessment. I doubt whether there has ever been so much interest at all levels in changing testing nor have the expectations for a new assessment system ever been higher.

Master teachers are able to see beyond the glamour of educational innovations. They pick and choose from new techniques and build a repertoire of educational resources. They don't overreact to new demands, but they don't ignore them either.

I hope we will use the same wisdom in approaching performance assessment — that in our urgency to make assessment right we proceed with reason and a clear understanding of what we are buying. I hope that assessment reform will benefit from all measurement strategies. I hope that we don't create unrealistic expectations. I hope that we'll hold on to our educational thermometers until we know we've got something better. ■

References

Torrance, H. P. 1991. *Can Measurement Really Drive Instruction?* The Case of National Assessment in England and Wales. Paper presented at the annual meeting of the American Educational Research Association. Chicago, Ill.

States Section. July 31,1991. *Education Week*, p. 2

California Department of Education. 1991. *A Sampler of Mathematics Assessment*. Sacramento, CA.

Richard Nicoll is director of research and evaluation for the Mt. Diablo Unified School District.

Reprinted with permission from MASC Journal, Spring 1990.

The Use and Abuse of Test Scores

What do the results really tell us?

by William J. Leary, Ed. D.

The use of test results by the public and news media has evoked considerable concern on the part of educators. Too often, test results among schools, school districts and even among students themselves, are utilized to formulate odious comparisons. The Secretary of Education's use, or misuse, of the "Wall Chart" is the latest example of these comparison-type techniques.

The use of test scores to compare performance levels among schools, school districts, or individual students or groups of students is questionable, at best. The public often uses test scores to determine how individual schools or school districts "compete" against each other. Local news stories either hail or decry each year's announced test results as a major indicator pointing to the success or failure of a given school or school district program.

If a district's test scores, whether SAT, ACT, or state-mandated tests, are up even slightly, success is proclaimed for the program. If the year's test scores are down, even slightly, administrators make haste to list all possible variables that might have occurred during the year to cause the "negative" results. Then, the school officials promise that the following year test results will improve based on local district or school initiatives instituted because of the "negative" test scores!

In fact, the taxpaying public has limited access to schools and also few formal ways to judge the performance of individual schools or school districts, other than the test scores. Only a small number of citizens, with the exception of some parents, and recently certain members of the business community, take the time or are given the opportunity to visit schools and observe what goes on in them. Certainly, some events are sponsored in which citizens are invited to visit schools or districts, but these projects are limited, and rarely allow citizens the opportunity to enact any kind of systematic evaluation of schools or a school district's overall performance. Thus, test results are used too frequently as a comparative tool. Instead, these scores should be better used as diagnostic tools, designed to assist teachers and students to assess their own performance and to grow from the proper analytic use of test results.

But what do our present testing methods really tell us about student progress? Carl Bereiter, in his article, *Using Tests to Measure Change* (1962), and writing about the commitment to education and how we use evaluation of educational practice, states "the only reasonable evaluation of an educational practice is one that measures . . . changeable traits." He goes on to point out that tests available for use in evaluation, "are designed as predictors of future status," and are therefore insensitive to the changes educators are trying to develop and discern.

Haywood et al, in their work, *Behavioral Assessment in Mental Retardation* (1975), indicate that tests traditionally determine intelligence by measuring achievement relative to a student's age. Therefore, these tests judge learning that took place in the past. The Haywood group also states, "Such an approach assumes that individuals being compared have had an equal opportunity to learn... [This] assumption is patently untenable."

J. Korp and I.E. Sigel, in their work, *Psychoeducational Appraisal of Disadvantaged Children*, also point to the effect that the environment plays on student development. "Scores on standardized individual and group intelligence tests have been known to vary according to the child's membership in a particular socioeconomic class, the area in which he lives, the level of education of his parents and his ethnic origin." Unfortunately, these very factors are often used, incorrectly, to indicate genetic factors as a reason for poor performance, which in turn justifies some forms of discrimination against certain students. Another incorrect use of tests has occured with the SAT exam. This test has become a barometer for our country's educational performance. Its original use was intended to be supplemental in nature, to assist colleges in making admissions decisions. Now, it is frequently utilized by colleges and universities as the ultimate criterion to exclude students from those institutions. When used with other criteria, such as class rank and extracurricular activities, SAT

If the year's test scores are down, even slightly, administrators make haste to list all possible variables that might have occurred to cause "negative" results.

William J. Leary is Superintendent of the Gloucester, MA Public Schools.

BEST COPY AVAILABLE

scores also can be used to include those students a university wishes to accept. But it does little-if anything-to predict student academic success in the first year of university study.

In other situations, tests and test makers are becoming much like the tail that wags the dog. As Samuel Meisels points out in *High-Stakes Testing* (1989), curricula decisions in some communities are being determined by what the test questions are, giving control of the curriculum to the test developers—far from the goal of helping teachers to determine where students may need help.

Interpreting test results has become mixed with political needs, some social requirements, and various theoretical concepts which can involve legislative policy. Thus bias enters into determining what test results mean. Value judgments by the press and the public often lead to inaccurate assumptions as to the meaning of interpretations of test scores. So, what is the public to do in order to better assess student progress?

Some solutions to this dilemma of student diagnosis and evaluation are available, though at a financial cost. Changes in how students are tested, class size, different school structures such as is advocated in Theodore Sizer's *Coalition of Essential Schools*, use of tutors, and peer learing groups are just some of the alternate strategies that might be used.

Also, Howard Gardner's work on the theory of multiple intelligences, summarized in his book *Frames of Mind* (1983), offers new ideas in developing programs to gauge how children learn developmentally. Gardner and his colleagues believe that current testing programs engage primarily two kinds of intelligence—linguistic and logical mathematical faculties. They want education to look at other additional abilities in children so that students are allowed to demonstrate their strengths.

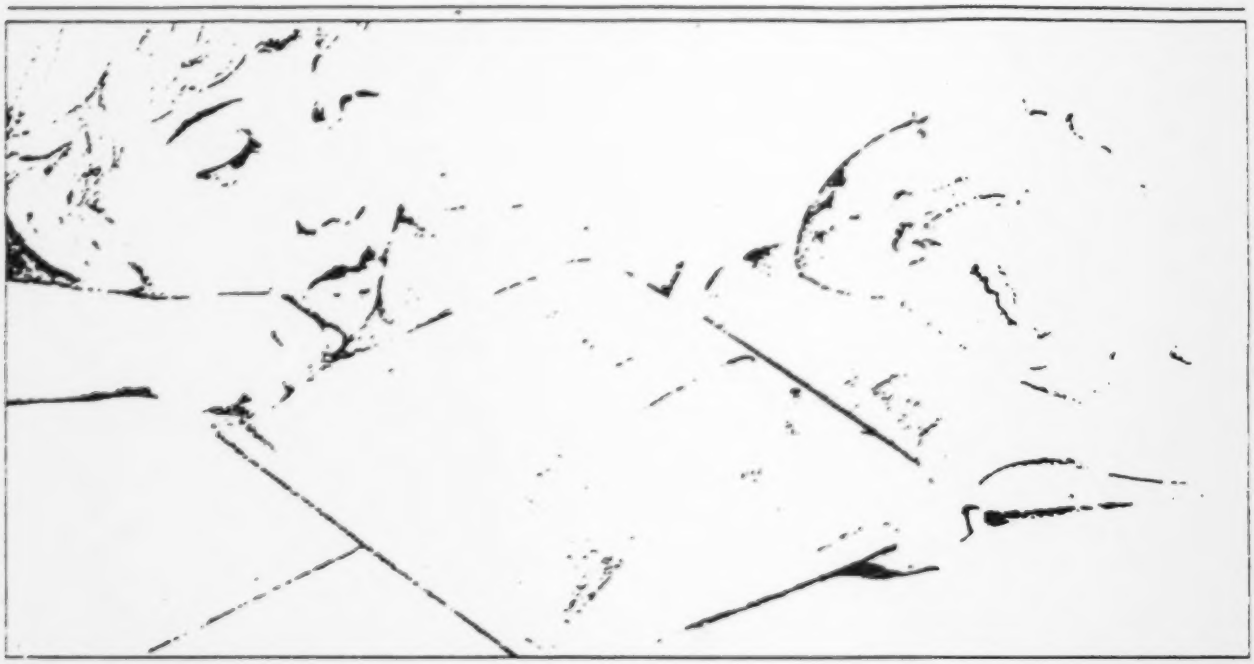
In other situations, tests and test makers are becoming much like the tail that wags the dog.

With these and other assessment ideas such as the "Arts Propel" project in Pittsburgh, where students compile portfolios as a major part of their grade, and the principle of "Diploma by Exhibition" concept espoused by the Coalition of Essential Schools project, provide alternatives to psychometric testing for determining student skill levels. In addition, these examples offer another method of informing the public about student progress and the quality of schools.

It is necessary that educators, students, the public, and the media have some standards by which to judge our schools. It is also time that all of these groups look at additional methods, other than test scores to make those judgments. □

Reprinted with permission from Phi Delta Kappan, February 1991.

Linking Restructuring to Authentic Student Achievement



Unless plans for new structures are guided by a vision of educational outcomes that articulates new standards for student achievement and for teaching, restructuring is not apt to produce favorable results for students, Mr. Newmann warns.

BY FRED M. NEWMANN

THE CALL to restructure education sends an ambiguous but serious message. The word *restructure* suggests that the system is so fundamentally flawed that such terms as *innovation* and *reform* aren't robust enough to describe the changes that are needed. But what exactly is the challenge that we face in restructuring? What needs to be changed, in what ways, and for what purposes?

The history of education in America has been punctuated repeatedly with powerful slogans that mobilize the energy of practitioners, researchers, and policy makers. Recent examples include "back to basics," "relevance," "community control," "cultural literacy," and "effective schools." Much of a slogan's appeal rests in its capacity to conjure up multiple meanings that can draw diverse constituencies together in a common cause. While a slogan galvanizes public opinion and focuses energy, thus offering new possibilities for action, its ambiguity brings the great risk that energy will be dissipated in scattered, even contradictory, directions.

Therefore, before committing ourselves to new structures, we should try hard to reach agreement on the nature of the sys-

FRED M. NEWMANN is a professor of curriculum and instruction at the University of Wisconsin, Madison, and director of the Center on Organization and Restructuring of Schools. *This article is adapted from a speech presented at the Indiana University Annual Education Conference, held in Bloomington in April 1990, and its writing was supported by Indiana University, by the Office of Educational Research and Improvement of the U.S. Department of Education, and by the Wisconsin Center for Education Research. However, the opinions expressed are the author's own.*

Illustration by Thomas Ries

tem's failure. For without agreement on the nature of the illness and on the purpose of treatment, the pursuit of radical cures could do more harm than good for students.

Today, reformers tend to address two main issues. The most obvious is the large proportion of students, especially low-income students of color, who fail in school and score poorly on nationally standardized tests. Much of the rhetoric about national education goals reflects concern about this group. Another issue – less popular, but perhaps even more disturbing – is that even those who succeed in school and score well on conventional tests have not been educated to cope successfully with the demands of personal, vocational, and civic life in contemporary society.[1]

Is the point of restructuring to provide a better way of teaching the current curriculum to students who haven't learned it through the old structure? Or is restructuring to be a vehicle to fundamentally change what is taught and how it is taught to all students? These alternative aims for restructuring can reflect conflicting goals of education, and the goals held by those pursuing each aim may lead to the creation of very different types of educational environments. It may not be necessary to choose between these stark alternatives, but the tension must be recognized. It reflects a persistent historical debate about educational ends that has no simple resolution and that even the restructuring movement will not solve once and for all.

Regardless of one's position on the forms of competence that schools should promote and regardless of the fact that large proportions of parents and educators perceive the schools to be working well, there are at least two grounds for seriously entertaining the prospect of fundamental organizational change. First, teachers face the persistent difficulty of engaging students in serious academic work in schools as we know them. Except for a few highly motivated students, most young people complete school only as a ritual. This pervasive disengagement creates massive problems of crowd control for educators and wastes the time of students and staff members alike.

Second, we must recognize the unfortunate record of traditional efforts to teach a virtually limitless list of fragmented pieces of declarative knowledge – e.g., that the issue of slavery divided the U.S. in the 19th century or that squares have equal sides and 90-degree angles. This enterprise, pursued over the ages as the main task of formal schooling, now actually undermines education. When schools try simultaneously to keep up with the explosion of knowledge and to accommodate the host of social demands placed on them, they cannot possibly give students the opportunity to develop in-depth understanding and the capacity for higher-order thinking that are required for success beyond school. The problems of disengagement and of excessive emphasis on coverage are related, and they exact their most tragic toll on students who, because of low income, cultural background, or lack of social support, do not succeed in schools as they are currently organized.

One of the biggest mistakes we make is the order in which we design the components of the education system. First, we set up an organizational structure; next, we plan the curriculum to fit the structure; and finally, we choose criteria for student success. Not until this last step do we articulate in a concrete way the ultimate educational purpose of the system. But by this point the outcomes are largely predetermined by the organizational structures and curriculum. Thus it should not be surprising that schooling seems so dysfunctional.

Moreover, the restructuring movement is about to perpetuate this error, because it, too, is trying to design organizational structures before clarifying purposes and reaching consensus on the educational ends that organizational structures should serve. For example, most of the talk one hears about school-site management, career ladders for teachers, or schools of choice never mentions how these mechanisms will teach students to write about literature, to reason about scientific phenomena, or to learn important geographic facts.

There is no reason to believe that restructuring will produce favorable results for students unless plans for new structures are guided by a vision of educational outcomes that articulates new standards for student achievement and for teaching. In this article I hope to nudge restructuring in this direction. First, I propose *authentic* achievement as the primary goal for students. Then, I advocate the need for *substantive conversation* in teaching, a new standard that is necessary to promote authentic student achievement. Finally, I suggest particular structures for student work that can facilitate authentic achievement and substantive conversation.

AUTHENTIC ACHIEVEMENT INVOLVES THE CHALLENGE OF PRODUCING, RATHER THAN REPRODUCING, KNOWLEDGE.

AUTHENTIC ACHIEVEMENT

The kinds of skills required to earn school credits, good grades, and high scores on typical tests are often considered trivial, meaningless, and contrived – by both students and adults. In contrast, a "restructured" vision of the goals of education seeks to evaluate performance activities that are worthwhile, significant, and meaningful: in short, activities that are authentic.

What criteria help us to recognize authentic forms of academic achievement? Consider the achievements of successful adults – scientists, musicians, business entrepreneurs, politicians, craftspeople, attorneys, novelists, physicians, designers, and so on. What are the characteristics of their work that justify calling their accomplishments authentic rather than contrived and trivial? Can we identify key distinctions between these authentic accomplishments and the work that students complete in school?[2]

People in the diverse fields named above face the primary challenge of producing, rather than reproducing, knowledge. This knowledge is expressed through discourse, through the creation

Accomplishments should have intrinsic value beyond their value in assessing knowledge.

of things, and through performances. We do not expect children to attain levels of competence comparable to those of skilled adults, but we do want students to develop in the same direction.

To progress on the journey toward adult competence, students should set their sights on authentic expressions of knowledge. That is, they must hone their skills through guided practice in producing original conversation and writing, through the repair and building of physical objects, through artistic and musical performance. The conventional curriculum asks students only to identify the discourse, things, and performances that others have produced – e.g., by recognizing the difference between verbs and nouns or between socialism and capitalism; by matching authors with their works; or by correctly labeling rocks and body parts. Certainly, the production of knowledge must be based on an understanding of prior knowledge, but the mere *re*production of that knowledge does not constitute authentic academic achievement.

A second defining feature of authentic academic achievement is its reliance on a particular type of cognitive work that can be described as "disciplined inquiry." Disciplined inquiry, in turn, seems to consist of three features: 1) use of a prior knowledge base, 2) in-depth understanding rather than superficial awareness, and 3) production of knowledge in an integrated (rather than fragmented) form. In highlighting these features, I am not suggesting that students should be expected to make seminal contributions to the academic disciplines, to the professions, or to the arts. However, I am suggesting that students are quite capable of engaging in these forms of cognitive work when they are adapted to the students' levels of development.

1. *Prior knowledge base.* For new knowledge to be significant and valid, it must be based on substantive and procedural knowledge that has been accumulated previously by workers in the field, who establish facts, vocabularies, concepts, theories, algorithms, and conventions for the conduct and expression of inquiry. However, relying on a prior knowledge base does not always mean agreeing with intellectual traditions or continuing to develop them in a linear fashion. The ultimate point of disciplined inquiry is to move beyond existing knowledge – through criticism and the development of new models that are themselves stimulated by prior knowledge. Most of the cognitive work of schools consists of transmitting prior knowledge to students and asking them to reproduce it, rather than helping them to use it to produce new knowledge.

2. *In-depth understanding.* Disciplined inquiry tries to develop in-depth understanding of a problem, rather than only a passing familiarity with or exposure to pieces of knowledge. One masters prior knowledge not primarily to become superficially knowledgeable about a broad range of topics, but in order to develop complex understanding of a smaller number of issues. Such detailed understanding is particularly necessary for the production of new knowledge. In contrast, many of the cognitive tasks of school ask students to show only superficial awareness of a vast number of topics.[3]

3. *Integration.* To produce knowledge, one must assemble and interpret information, formulate ideas, and make critiques that cannot be easily retrieved from the existing knowledge base. All these activities require the ability to organize, synthesize, and integrate information in new ways. Students are unlikely to succeed at such tasks unless they learn to look for, to test, and to create relationships among pieces of knowledge that otherwise appear unconnected.

The third and final distinction between authentic achievements and traditional school achievements is that authentic achievements have aesthetic, utilitarian, or personal value apart from their value in documenting the competence of the learner. Indeed, this distinction may be the most crucial. When people write letters, news articles, insurance claims, or poems; when they speak a foreign language; when they develop blueprints; when they create a painting, compose a piece of music, or build a stereo cabinet, they try to communicate ideas, to produce a product, or to make an impression on others beyond the simple demonstration that they are competent. Achievements of this sort have special value that is missing from tasks contrived only for the purpose of assessing knowledge, such as spelling quizzes, laboratory exercises, or typical final exams. In many cases, the cry for "relevance" is simply a less precise expression of the desire that accomplishments should have intrinsic value beyond their value as indicators of success in school.

To summarize, the idea of authentic achievement requires students to engage in disciplined inquiry to produce knowledge that has value in their lives beyond simply proving their competence in school. Mastery of this sort is unlikely to be demonstrated in familiar testing and grading exercises. Instead, such mastery is more often expressed in the completion of long-term projects that result in the creation of discourse, things, and performances of interest to students, their peers, and the public at large.

Why should we aim toward authentic achievement? I suggest at least two compelling reasons, which respond directly to the problems of the conventional curriculum that I mentioned earlier. First, participation in authentic tasks is more likely to motivate students and to sustain the hard work that learning requires. Because authentic work has value beyond the mere demonstration of competence in school and because it permits more comprehensive use of the mind, students will have a greater stake in authentic achievement. Second, authentic academic challenges are more likely to cultivate the higher-order thinking and problem-solving capacities that are useful both to individuals and to society. The mastery gained in school would then transfer more readily to life beyond school and thus increase the efficiency of our public investment in schooling.

To develop organizational structures consistent with this vision, educators will

BEST COPY AVAILABLE

need to plan "backwards." Beginning with a conception of authentic achievement and of the specific projects that express it, they will need to select appropriate content within the subjects. They will also need to consider the kinds of human interaction – involving teachers, students, and others – that promote disciplined inquiry into the subject. Although I cannot address organizational needs for specific subjects in this article, one aspect of interaction seems critical to authentic achievement in many subjects. This is the need for substantive conversation in instruction, which constitutes a major departure from present practice.

SUBSTANTIVE CONVERSATION

When the curriculum is geared toward the goal of authentic achievement, new forms of teaching will be necessary. No longer will it be possible for students to succeed in school simply by listening to the teacher and responding to questions that can be answered in half a dozen words. To complete authentic projects, students will need to formulate and to ask questions, to explain themselves to peers and to adults, and to refine their ideas. While many authentic achievements require capacities other than verbal discourse (e.g., working with one's hands, making mathematical estimations, using aesthetic sensitivities), substantive conversation between the learner and some supportive resource – teacher, peer critic, or other knowledgeable authority – is usually necessary. I emphasize substantive conversation, not because it alone is sufficient to develop all forms of authentic achievement, but because it is usually critical for success in authentic tasks and because conversation is the primary medium through which teachers and students communicate.

Substantive conversation differs from conventional classroom talk, in which the teacher usually has only two purposes (aside from disciplinary ones) in speaking to students: to transmit an item of declarative knowledge (usually a definition or a fact) or to determine whether the student can reproduce that item.[4] The student's goal (assuming that he or she desires to cooperate) is to give the answer that the teacher has in mind. The teacher maintains control in that he or she knows, prior to the interaction, what pieces of knowledge must be communicated and what answers will be acceptable.

In a substantive conversation, the purpose of talking is quite different. Each person is trying to express a point of view (often in order to persuade the other), to understand why the other holds a particular viewpoint or interpretation, or to arrive at a solution to a problem that neither has previously solved. This last purpose is the most difficult to fulfill, but it is perhaps the most likely to render the work authentic.

In pursuing any of these ends, the teacher uses talk to integrate subject matter into the student's meaning system, and the student tries to understand how the teacher's messages might empower him or her in the world. In this sense, each is interested in what the other has to say, the course of the conversation is less predictable than in conventional classroom exchanges, and each participant has some sense of ownership. To fulfill these purposes, substantive conversation requires sustained, continuous talk between two or more people. In school as we know it, the two key participants are usually student and teacher, but students can also learn through conversations with their peers, with other adults, and perhaps with well-programmed computers. Teacher/student conversation should remain central. But to expand students' opportunities for expression and feedback, it will be necessary to rely substantially on and to support these other sources as well.

WHEN THE CURRICULUM IS GEARED TOWARD THE GOAL OF AUTHENTIC ACHIEVEMENT, NEW FORMS OF TEACHING WILL BE NECESSARY.

What is the connection between substantive conversation and authentic achievement? Begin with the definition: authentic achievement is expressed in discourse, things, and performances that make reasonably complete, integrated statements and that reflect students' production of in-depth knowledge. Then ask the question, How can we instruct students to make such statements? We can show them models of authentic accomplishments of others, and we can give them critical elements of declarative knowledge on which to build. But how will they learn to make their own statements? Substantive conversation is the key. It provides the crucible for practice, for seeking new knowledge that relates to the problem at hand, and for trial, feedback, and revision. In short, substantive conversation forces us to transform declarative knowledge into applied, integrated knowledge.

STRUCTURES THAT FACILITATE AUTHENTIC ACHIEVEMENT

The Council of Chief State School Officers divides restructuring into four main categories: school governance (including decentralized authority and school choice), reforming the nature and organization of curriculum and instruction, new professional roles for educators, and accountability (especially new methods of assessment and direct state intervention).[5]

This analysis applies most directly to curriculum, instruction, and accountability. But I believe that, unless the entire restructuring agenda focuses more closely on authentic student achievement and on substantive conversation to attain it, restructuring will produce at best only cosmetic changes in students' education.

Assuming a commitment to these goals, what structural conditions seem necessary to achieve them? The conditions I propose below are in no way exhaustive. They outline certain key conditions under which students would work, but they do not provide structures of school governance, detailed systems of assessment, or specific new structures for the professional development of teachers. My argument has implications for these areas as well, and I do support the need for pro-

viding teachers with working conditions that enable new forms of professional support. But my central concern here is the structure of the instructional environment.

At least four conditions seem essential for students to complete authentic accomplishments: collaboration, access to tools and resources, worker discretion or opportunity for ownership, and flexible use of time.

Collaboration. Achievements outside of school often depend on the opportunity to ask questions of, to receive feedback from, and to count on the help of others, including peers and authorities. In contrast, typical school activities require each student to work alone. Working together is often prohibited, because it is seen as a form of cheating. Of course, it is important for students to learn to work on their own and not become too dependent on others. But if opportunities to cooperate are consistently denied, students will not experience a critical process that adults, both expert and novice, consistently rely on for success.

Useful collaboration depends to a great extent on opportunities for substantive conversation, which itself requires special structural conditions. More time will be needed for teachers to communicate with individual students through sustained talk and writing and for students to talk with one another. Substantive conversation also entails major shifts in the roles of teachers and students. Teachers will function more as mentors and coaches, less as depositories of static knowledge to be reproduced. Students will function more as constructors and producers of knowledge. They will rely on teachers for help, but they will not be mere absorbers or consumers of everything the teacher says. Students will also have to take on the new roles of seeking help from and giving help to one another as they learn.

Access to tools and resources. When competent people complete significant accomplishments – e.g., writing books, designing buildings, composing music, or reaching collective bargaining agreements – they usually have made use of a variety of printed materials and such machines as telephones, computers, and laboratory equipment. The information they require is generally too vast and too rapidly updated to commit to memory. In medieval times, before the invention of the printing press, becoming educated required memorizing the contents of key manuscripts, because future access to them was unlikely. But today increasing specialization and the explosion of knowledge create a new challenge. Possession of prior knowledge is critical to disciplined inquiry, but it is now impossible to teach students all the relevant information they will need to complete authentic tasks. The new challenge is to teach students to use tools and resources to find the appropriate knowledge when they need it.

Discretion and ownership. Rather than always toiling within the confines of predetermined routines arbitrarily dictated by authorities, the creators of authentic work influence the conception, execution, and evaluation of the work itself. At a minimum this entails exercising some control over the pace and procedures of learning; over opportunities to ask questions and to study topics deemed important; and over constructing and producing knowledge in one's own language, rather than merely reproducing the language of others.

There are, of course, important limits on the extent to which students should control their learning of academic subjects. Certain facts, definitions, concepts, algorithms, and processes of verification must be assimilated according to the standards of the disciplinary fields. But for this kind of learning to be translated into authentic products, students must enjoy some autonomy and discretion in its application.

Flexible use of time. The significant achievements of disciplined inquiry often cannot be produced within rigidly specified time periods. Adults working to solve complicated problems, to compose effective discourse, or to design products are rarely forced to work within the rigid time constraints imposed on students, such as the 50-minute class or the two-hour examination.

Standard time schedules stem from bureaucratic procedures designed to manage masses of students and diverse course offerings. Since they have nothing to do with the time requirements of disciplined inquiry, they can reduce the authenticity of student achievement. Achievements in noninstructional tasks, such as journalistic writing, interior design, or medical care, do involve deadlines and time limits. But in such work the schedules tend to be determined more by the nature of the work than by the requirements of institutional management.

The four conditions outlined above

WE MUST RESIST THE TENDENCY TO INSTITUTE ADMINISTRATIVE CHANGES DEVOID OF FRESH EDUCATIONAL VISION.

suggest at least the following changes in the structuring of students' work.

- To stimulate collaboration and the substantive conversation so crucial to it, more time must be allocated to individualized and small-group instruction.
- Students' access to knowledge must be enhanced by greater use of technology (telephones as well as computers) and by opportunities to learn from sources outside of school.
- Students need to exercise discretion and assume responsibility in the planning, execution, and evaluation of their work, which further underscores the need for individualized and small-group work along with assessment that accepts diverse ways of demonstrating competence.
- Instructional time should be organized to permit more sustained, long-term, and in-depth investigation, in contrast to the fixed time slots designed for survey coverage.
- Assessment tasks calling for the production of discourse, things, and performances will require new procedures and in some cases the participation of individuals who are not members of the school staff to evaluate student performance.

Instituting these changes will contribute to authentic achievement. But they alone cannot guarantee it, for ultimately the quality of student achievement will depend on the quality of conversation that students have with teachers and with one another. For years substantive conversation has been difficult to find in schools, but only partly because of structural constraints. Even if teachers were blessed with more time and much lighter teaching loads, I suspect that substantive conversation would still be rare, because the goals of authentic achievement and substantive conversation also presume major shifts in prevailing conceptions of education. Organizational changes alone will not modify long-standing conceptions of knowledge and deeply socialized habits of didactic teaching that direct most of our educational efforts toward inauthentic forms of mastery.

Just as plans for restructuring should begin with the conditions in which students learn, the next step is to consider the working conditions that teachers need in order to generate substantive conversation. Again, we must not expect organizational structures alone to transform teaching. However, given the complexities of the teaching task, we cannot emphasize enough the need to provide teachers and administrators with sufficient authority, time, and assistance to reflect critically on their practice, to experiment, to fail, and to try again. Beyond enabling teachers to increase the proportion of time they spend interacting with individual students and small groups, we must work to give teachers more opportunity to collaborate, and we must provide for regular, sustained professional development that focuses on the goals of substantive conversation and authentic achievement.

Experiences in restructuring to date reveal three obstacles that dramatize the need for this kind of support. First, teachers, parents, and students who have experienced only the conventional version of education – especially those who have been successful – cling tenaciously to it, even when they have the opportunity to make substantial changes. Faith in conventional practice is fortified by the large proportion of students who do succeed on conventional measures, both in school and beyond. Second, external pressures for district, state, and national accountability, coupled with the need for ways to determine admission to college, continue to sanctify mastery of declarative knowledge as the most important goal of schooling. Finally, the struggle to establish new collaborative roles for teachers is stressful and time consuming; it drains energy and leaves diminished personal resources for thinking about new forms of achievement and curriculum.

The history of education reform suggests that we have yet to see fundamental change on a grand scale.[6] We might argue that the contemporary sense of urgency and the scope of the current restructuring movement combine to offer a unique opportunity. But to avoid the costly mistakes of previous efforts, we must resist the tendency to institute administrative changes devoid of fresh educational vision. If we keep our sights on the targets of authentic student achievement and substantive instructional conversation, restructuring could form a new chapter in educational history – and one well worth the effort.

---

1. Albert Shanker, "The End of the Traditional Model of Schooling – and a Proposal for Using Incentives to Restructure Our Public Schools," *Phi Delta Kappan*, January 1990, pp. 344-57.
2. This complicated matter involving controversial educational values has not yet received extensive scholarly attention. Yet it begs for interdisciplinary analysis of the relationship between formal attempts to educate and the ways that human competence is expressed in noneducational settings. The conception I propose here builds on Harold Berlak, ed., *Toward a New Science of Educational Testing and Assessment* (Albany: State University of New York Press, forthcoming); Fred M. Newmann and Doug Archbald, *Beyond Standardized Testing: Assessing Authentic Academic Achievement in the Secondary School* (Reston, Va.: National Association of Secondary School Principals, 1988); John Raven, "Toward a Model of Competence, Motivation, and Behavior and a Paradigm for Assessment," in Berlak, op. cit.; Lauren B. Resnick, "Learning In School and Out," *Educational Researcher*, December 1987, pp. 13-20; and Grant Wiggins, "A True Test: Toward More Authentic and Equitable Assessment," *Phi Delta Kappan*, May 1989, pp. 709-13.
3. A commitment to depth over coverage entails no retreat from liberal education into narrow forms of vocationalism, as some might expect. The effort to concentrate on complex understanding welcomes study of a wide variety of disciplines and paths to knowledge. Reducing the emphasis on superficial exposure implies no necessary narrowing of the curriculum as a whole.
4. This analysis is based in part on the work of Martin Nystrand and Adam Gamoran, "Student Engagement: When Recitation Becomes Conversation," in Hersholt Waxman and Herbert Walberg, eds., *Contemporary Research on Teaching* (Berkeley, Calif.: McCutchan, forthcoming).
5. *Success for All in a New Century: A Report by the Council of Chief State School Officers on Restructuring Education* (Washington, D.C.: Council of Chief State School Officers, 1989).
6. Larry Cuban, "Reforming Education, Again, Again, and Again," *Educational Researcher*, January/February 1990, pp. 3-13.

Reprinted with permission from Phi Delta Kappan, February 1989.

How Can a Student 'Fail' To Be Educated?

Students are in school to obtain useful knowledge and to develop their natural strengths, Mr. Rogers maintains, but grades and competition do not contribute to these goals.

BY PHIL ROGERS

THE FACT IS, a student can't "fail" to be educated. It is only our educational system that has adopted the term *fail*. When an infant reaches for an object that has aroused his or her curiosity, do we say, "You have five tries, and if you don't grasp the object you have failed"? No. We might push the object closer. We might say things that the infant understands only in some vague but positive way, such as "Wow! That was close! Come on, you can do it! *Very* good!" The infant is already motivated by color, shape, and a desire to investigate the object. We provide encouragement.

We're not disappointed if the infant does not succeed. We don't show disapproval. The infant hasn't failed. We know that there will be another time and that the infant will succeed. Infants are born with a curiosity about their environment and a desire to satisfy that curiosity, a desire to learn. One of the purposes of education is to encourage, guide, and facilitate the satisfaction of inherent curiosity.

Eventually the infant matures enough to attend school. Given new material to learn, that child is now told that he or she must succeed in learning it in a certain length of time. If the student doesn't learn the material in that time, he or she is "failed," with all the attendant stigma. A failing grade is duly recorded on the student's record, where it generally remains "forever." The student is a "failure." What happened to the limitless tolerance, understanding, and encouragement that were shown when the child was an infant? They have disappeared. The student has entered the competitive and somewhat brutal world of "education," which may continue through college.

Does the term *failure* mean that the person is a failure? No. It simply means that the person did not meet minimum standards that indicate mastery of certain subject matter. The term *failure* is an admission that we are not all interested in the same things and/or that we do not all learn the same subject matter at the same age. Is that bad? No. It is normal.

Within the educational system, however, the term unofficially places a person in a negative position relative to other, "successful" individuals. It attacks the person, not the lack of mastery of some skill or knowledge. Does such a term belong in our schools?

Before examining in detail the changes that are needed, let us ponder another word – *competition* – that has an unofficial place in our educational system. Our grading system finely divides and categorizes students, setting up competition among them for a limited number of "good" grades. Does such competition belong in an educational system? Emphatically not! Does *any* competition belong in an educational system? Yes. Competition with oneself can be constructive. But any other kind of competition can be destructive: of relationships with other people; of sharing in the learning process; of physical and mental health; of feelings of self-worth, self-confidence, and self-esteem. Competition among students for grades contributes to the concept of "failure" in our educational system, since, by its very nature, competition produces winners and losers (i.e., "failures").

Not all competition with others is bad, of course. Competition in sports can produce better athletes and encourage teamwork, for example. In the business world, competition among products can produce better and less-expensive products.

Besides competition, another contributor to failure of the individual in the educational system is a sometimes unenthusiastic and disruptive learning environment. Such an environment can be attributed in part to our national commitment to "mass education" from kindergarten through high school. Mass education often becomes a babysitting service or a detention system. A small minority of students who are not ready or who are unwilling to be in the system may prevent or diminish the learning experience for the rest of the students.

To eliminate such conditions and to build a functional "mass education" system, minimum criteria for behavior, appearance, and performance must be developed and enforced. Stated simply, students should have consideration for one another, for the school staff, and for the school facilities. They should take some responsibility for promoting and maintaining the learning environment. The educational system should then be open to all who subscribe to these minimum criteria; those who do not subscribe

PHIL ROGERS is a professor of engineering at Saddleback College, Mission Viejo, California.

> Students who concentrate on areas in which they excel are more likely to develop self-confidence and self-esteem.

should be temporarily exluded from the mainstream and channeled into a speçial detention/rehabilitation section of the educational system until they are ready and willing to adhere to these criteria.

There should be no stigma attached to temporary exclusion from the mainstream. The attitude should be, "We will welcome you into our educational system, when you want to join it and are ready to do so, as evidenced by your subscribing to our criteria – and the sooner, the better." Of course, eliminating the concept of "failure" and reducing competition among students should increase their success rate and thus their feelings of self-esteem. In such an environment the number of unruly students should decrease. And the ways in which these changes would transform the educational environment should increase all students' chances to succeed.

If the educational system is to cease damaging individuals, other changes are needed as well. One area that requires attention is the grading system. Colleges, graduate schools, and employers do not need to know whether a person is an "A," a "B," or a "C" student, in order to accept or to hire that person. As a matter of fact, brilliant students with top grades are not necessarily the most productive and creative employees. However, schools and employers do need to know that a person has mastered certain subject matter and can use that knowledge.

I propose a three-tiered grading system from kindergarten through graduate school. It would include a grade (perhaps "E," for *excellent* or *exceptional*) to recognize outstanding or brilliant students. It would include a grade (perhaps "T," for *try again*) for those students who do not meet the minimum standards in a given subject. The third grade (perhaps "M") would indicate mastery of the subject matter, demonstrated by having met the minimum standards. Most of the students studying a particular subject would probably receive grades of "M." Categorization should not be more specific than this, since no one needs to know more about a person under consideration for further schooling or for a job.

Where would the standards come from for this proposed three-tiered grading system? Standards already exist for each subject taught. For use in the three-tiered grading system that I have proposed, these standards would only have to be reexamined and redefined to be sure that they are sufficiently rigorous.

Indeed, the users of grades, such as schools and employers, do not need to know whether a student received an "E" or an "M." Thus a two-tiered grading system – employing only grades of "M" and "T" – is possible. In fact, all that the users of grades really need to know is that a student has met minimum standards for the subjects taken. Therefore, a one-grade system – employing only the grade of "M" – would suffice. Such a grading system would eliminate competition among students for grades and eliminate the whole notion of "failure," since no grade would be recorded on a student's record until that student had successfully completed a course, thus earning an "M."

Next, I propose a change in attitude. There is no failure in education! Even in a three-tiered system, "T" is not a failing grade. It simply says, "Hey, you haven't finished this subject yet. Come on back and try again when you want to and are ready." There should be no stigma attached to the grade. After all, the student may not be curious about, interested in, naturally suited for, or mature enough to handle the particular subject area – but he or she may be quite competent in some other subject area.

Ideally, the educational system should endeavor to find and develop each individual's strengths, rather than "hammer away" at the individual's academic weaknesses. Its primary focus should be to determine as early as possible what subjects come easiest for students, give them the most pleasure, and lead to the highest levels of achievement. Then the system should strengthen and build on those subjects.

Additionally, the educational system should help the individual to meet minimum standards in other basic subjects that are necessary for survival and effective functioning in society. Students who concentrate their efforts in areas in which they excel and find pleasure are more likely to develop the self-confidence and self-esteem that come with success. Students are in school to obtain useful knowledge and to develop their natural strengths, in order to use them in their own lives and for the betterment of society. Grades and competition do not contribute to these goals. ■

SECTION 2

CLASSROOMS AND SUBJECT AREAS

Reprinted with permission from Educational Leadership, March 1990.

Finding the Value in Evaluation: Self-Assessment in a Middle School Classroom

Portfolios allow teachers to get to know their students—as readers, writers, thinkers, and as human beings.

. . . We must constantly remind ourselves that the ultimate purpose of evaluation is to enable students to evaluate themselves.
—Arthur L. Costa (1989)

Sarah was adamant. "They don't know me as a person and a writer. They don't know how I've improved." I watched as Sarah read the one mark on her writing sample—a "7." On a test mandated by our school district, Sarah had received a 7 out of 8, certainly a good score. But it didn't matter. "What does this tell me?" she continued. "I'm one less than an eight and one more than a six. So what?"

Sarah was right. The writing sample didn't show who she was as a person or a writer. And the response she received didn't help her. No one who had read her piece knew where she'd been, so how could any one tell how much she had grown? How sad, I thought, especially when all the evidence was right here in the classroom.

A Wealth of Information

The evidence was in Sarah's portfolio. In my classroom, portfolios have become each student's story of who they are as readers and writers, rich with the evidence of what they are able to do and how they are able to do it. Each portfolio is a collection of each student's best work.

I impose the *external* criteria for the portfolios—each student's two best pieces chosen during a six-week period from his or her working folder, trimester self-evaluations of process and product, and, at year's end, a reading/writing project. The students determine the *internal* criteria—which pieces, for their own reasons. I invite them to work on reading and writing from other disciplines and to include them in their portfolios, if they think the efforts are some of their best. Here are some examples. Joel's portfolio has a piece entitled "The King and His Achievements," a paper he wrote for social studies but worked on in a remedial reading class and in English. Sara and Jennifer each included a children's book they wrote collaboratively about a little boy investigating tidal pools along the coast of Maine. It was written for science, based on their field trips, as part of their study of marine biology. In English class Jen wrote, while Sara, studying the work of Trina Schart Hyman and Jan Brett, drew the illustrations.

Teacher as Learner

My classroom was not always like this. It has evolved slowly. I used to make all the decisions about what the students read and wrote and what they learned from that reading and writing. Then I would test them on all the information.

But times have changed. I've been turning the responsibility for learning over to them—they choose what they write, what they read, and what they

Photographs by James Whitney

Two students at Oyster River Middle School in Durham, New Hampshire, discuss how to arrange their writing from most to least effective in their portfolios.

need to work on in order to get better at both. I invite them to try different genres of writing, and I share a variety of literature that I love with them. They used to keep writing folders and were judged on all their writing. Now *they* select the best pieces to revise and rework. The portfolio of best pieces is separate from the working folder.

My students, however, aren't the only ones who keep portfolios. I also keep a portfolio. My reasoning is that if I don't value what I ask my students to do, then they probably won't value it either. What's in my portfolio? An education article entitled "Seeking Diversity," a poem, a personal narrative about my mother's sewing, and a letter written to the governor nominating our parent group for a state award for support of our Arts in Education program. I begin my writing and reading with the students. I share my drafts in progress. If I'm going to trust and value my students as learners, I have to trust my own possibilities as a learner.

Evaluation, the Last Holdout

For years I was the final decision maker about how well each student did. But then I began to wonder what would happen if I considered my students the best evaluators of their own writing and reading, both in progress and as a final product. I decided to find out. I not only let them choose their own topics, but I let them choose the pieces that were going best. So that they would have a selection to choose from, I asked them to write at least five rough draft pages a week. Further, I not only let them choose their own books, but I let them select the ones they wanted to respond or react to.

Over time, I began to see more diversity and depth to their writing, their reading, and their responses to literature. I discovered that the students knew themselves as learners

I impose the *external* criteria for the portfolios, and my students determine the *internal* criteria.

As students learn to critique their works, their abilities as writers and readers often deepen in the process.

better than anyone else. They set goals for themselves and judged how well they had reached those goals. They thoughtfully and honestly evaluated their own learning with far more detail and introspection than I thought possible. Ultimately, they showed me who they were as readers, writers, thinkers, and human beings.

As teachers/learners, we have to believe in the possibilities of our students by trusting them to show us what they know and valuing what they are able to do with that knowledge. The process of turning learning back to the students, from choice of topic or book all the way through to the evaluation of their own processes/products, produces students like Nahanni. I chose Nahanni, not because she had the best portfolio, not because she was the most articulate, but because she was a little better able to reflect on what she did. Through Nahanni's portfolio I want to show the possibilities in diversity, depth, growth, and self-evaluation. This kind of evidence shows the value in evaluation.

One Student's Story

Nahanni's portfolio is typical of the diversity and depth of writing I am now seeing in my students. She has 10 final drafts: 3 poems, a personal narrative, a character sketch, a letter, a pen and ink drawing, a play, an essay, and a picture of her final project for the year, an acrylic collage representing her interpretation of the book *Night* by Elie Wiesel. Everything that contributed to the final draft—lists, rough drafts, sketches of ideas—is attached to the final draft of each piece. Nahanni's evaluations of herself as a writer and reader are also in the portfolio, just as they are for all my students.

Nahanni ranked the following poem as her best piece for the year.

Looking across rows of music stands
I see an ordinary face—
a quarter note on a sheet of paper.
I know the face has laughed and cheered,
and laughed some more,
and cried.
Violins raise, bows sway, and the ordinary quarter note frowns and comes alive.
The lone quarter note is gently carried off the page and lightly danced onto the air.
Hidden behind Beethoven's Fourth, one little
second plays along.
A royal theme, issued from the staff, fades, then crescendos to a wail.
My loving ear fills up with sounds of staccato arpeggios, an Allegro, and a triste.
And so, to my dear little—quote—unimportant second,
I give my ballad of soft support.
Bring your laughter and your cheers, your sorrows, and your fears,
We'll laugh together, heal the pains, use up some tears.
We'll share in both.

Nahanni reflected on her poem and why she chose it:

My best piece is my best . . . because it meant the most to me. I used music to portray a friendship. I think there are layers of understanding, and to enjoy the poem you don't need to understand all the layers.

The piece that is the least effective means nothing to me. It was an assignment. That's why the pieces that are my best are so important. I chose topics that were important to me.

Now I know that in order to write something well, you have to care about it. The first important thing is that you like a piece of writing, then you worry if anyone else likes it . . . I've learned to add detail to get something across. I've learned to care about my writing and write in order to resolve things, because even when I read a piece over, I always learn something.

I think writing is like visiting places, you see different things each time. You read a piece of writing and you think you read it and saw it, but then you go back and read it again and see new things . . .

Writing isn't just a school subject. It's part of how we think. I write because I need to figure out what I'm thinking.

Nahanni's final version of the poem, though, is a far cry from her first draft, which looked like this:

Letter to my best friend—You are my very best friend and I think you know I don't mean that in what you'd call a "superficial" way . . . It feels like we are always one person. I want you to know how much I care about you. Never forget me, I will NEVER forget you. The times with you will never leave me.

How did she get from this early draft to a quite different final version? For some insight, let's look at Nahanni's second draft, exactly as she wrote it:

Two worlds collided and they could never ever tear us apart.
Two hearts that can beat as one, there aren't a single thing we can't overcome.
We're indestructable.
The light of a billion stars pales
As we comfort each other . . .
. . . the universe and a billion stars—and you and me.
Crowded there with others or alone.
At times it seems—we will always be together
So bring your laughter and your cheers your sorrows, and your fears.
We'll heal the pains, use up some tears, we'll share in both.
poem or letter? [She asks herself.]

Nahanni explained that her first breakthrough on this piece came when she answered my question, "Where did you first think of writing this to your friend?" She said she looked up across rows of music stands and saw her friend—this ordinary

face. I asked, "What's an ordinary face look like?"

She continued: "Like . . . like . . . a quarter note on a sheet of paper . . . just an ordinary note." And Nahanni was off. She made lists of musical terms. Played with them in phrases. Literally cut the phrases apart and made two piles—the ones that related best to what she was trying to say about friendship and the ones that didn't feel right to her. She read drafts in progress to her peers and to me for responses to what was working, to hear the questions we had, and to get suggestions.

I prompted her to elaborate on the poem's development:

In my first draft I didn't tie in much about music to my theme of a person—or face. I made a quarter note be the face. I use a piece of music to interpret what the face fears and what I feel about that face. I feel now that what I have to say is more important than trying to shape and twist the words into specific phrases and lines. This is one of my best pieces because I've worked so hard, making all the words say exactly what I want to mean.

At the beginning of the year, Nahanni had set three goals for herself: to try writing poems (which, as we have seen, she has done), to write longer pieces, and to send a piece off for consideration for publication. To meet her second goal, she wrote a prose piece entitled "Melted Wax." At first, she had steered away from writing long pieces because she was "afraid of losing peoples' focus, or maybe I'm afraid I'll lose my own focus." She discussed how the piece came together:

I was brainstorming all the positives and negatives in my life. . . . trying to think of a way to write down what I thought of Linda's death. One of the things that makes the piece good is the layers of depth that people can find. Most people can come up with the point of the candles . . . but if they don't, they can still enjoy it. I changed the lead all around. My first lead was dull and didn't say much. Now it starts with a quote and pulls people in better.

Nahanni also accomplished her third goal: to send a piece for publication. The poem she selected had gone through as many changes as her first poem.

My students choose what they write, what they read, and what they need to work on in order to get better at both.

I like this poem, but it doesn't have that much meaning to me. I sent it to *Merlyn's Pen*, and that's what they told me. I knew they would. I think I'll make it how they'd like it since if it's changed, it's not a problem for me because I don't think it's perfect the way it is. It's not for me that I want to keep it that way. Like the poem I wrote for my friend, "Looking Across Rows of Music Stands." I wouldn't change that no matter what anybody thinks. It's for me—and my friend.

This year Nahanni started reading 30 books—and finished 29. Her favorites were: *The Little Prince* (de Saint Exupery), *One Child* (Hayden), *The Chocolate War* (Cormier), *Beyond the Chocolate War* (Cormier), *Night* (Wiesel), and *The Princess Bride* (Goldman). She has formed some definite opinions on the subject of reading:

In order to be a good reader, I think you must read a lot and think about what you are reading—how it relates to you, what the writer wants you to think versus what you really get out of it. The responses I got to my reactions to the books were written or asked in such a way that I feel my ideas are important, and that makes me think more.

I wanted to read faster . . . I don't think I accomplished this, except that now I read more than I did before. I don't think [reading faster] was a really important goal. I wanted to understand more, and I think I've done that.

Nahanni believes that if a book is really good, you can learn a lot about what the author thinks about life. In a good book, the meaning is made clear so that "I think I've discovered it on my own, not so it was told to me."

Nahanni doesn't think she consciously connects reading and writing, but listen to what she consciously does:

I think I've been looking a lot more for metaphors and hidden meanings in my reading. This is a result of my change in writing—or vice-versa. I think that now I look for things that aren't so obvious. I've discovered that those things mean so much to me. The words that are unwritten teach much stronger than words on paper.

Self-Evaluation, Self-Discovery

I don't have to be the sole evaluator of Nahanni's writing and reading. She's far better at it than I am. And the better she knows her own process as a writer and reader, the better she becomes at both. But I *can* evaluate Nahanni's growth as a writer and reader if I have to. I have all the evidence in front of me in her drafts of writing from rough to final, in her responses and reactions to what she's reading, and in her self-evaluations of herself as a writer and reader. As teachers, we must listen first to the perceptions our students have of themselves—and address what they think they can and cannot do. From what Nahanni showed me, I wrote the following narrative about her growth as a learner:

Nahanni has become an independent learner who reads and writes for her purposes because she wants to become better at what she does. Her reading and writing show me she is engaged in the excitement of learning. She is a keen observer of life around her, often finding topics because she is always looking. She has an acute sense of detail, able to see, hear, smell, touch, and feel things that many of us miss, all with a sensitivity to the human factor. She is a caring, sensitive young woman who leaves an impact.

As a writer Nahanni knows how to find a topic, how to play with words until they say exactly what she means, how to seek help for revision, how to use a variety of resources, how to ask questions, and how to answer them. She is not afraid to take risks with new genres. She has tried poetry this year, while still attempting a variety of prose pieces. There is always a meaning to her writing—a reason to read it. She has a message for her reader, because the message is always for her, first.

Nahanni has an acute sense of audience. After receiving a rejection from one magazine I asked her if she wanted to revise the piece and resubmit it. "No," she said, "this wasn't for the magazine anyway." Another piece she will revise because it wasn't for her alone.

BEST COPY AVAILABLE

Her images are fresh and vivid. She searches for precise words. Her dedication is evident in her willingness to revise until pieces say exactly what she means, in the appropriate format. She grapples with big issues: friendship, control, love, prejudice, hate, fear, uniqueness.

She is a thoughtful reader who reads a variety of books for a variety of reasons: some to make her think, others just for fun. She knows how to take meaning to and meaning from a book. She relates her own life experiences to the experiences she draws from books. She comprehends, analyzes, reflects, contrasts, compares, synthesizes, wonders, questions, and criticizes literature.

I believe her writing has changed her perceptions as a reader. At first she just wondered why authors did certain things, now she reflects on how she would have done it differently. Nahanni writes and reads to find out what she is thinking.

What Can Teachers Do?

Will the portfolio concept work in every classroom? Yes, but certain conditions must be present. First, students must be immersed in reading, writing, speaking, and listening. Second, they need to be given time to do so in large blocks. Third, they must be given choices about what they are doing. And, fourth, they must receive positive response to their ideas.

Once these conditions exist, teachers can introduce the concept of portfolios: as places where students collect evidence of who they are as readers and writers. It's also a great idea to keep and share your own portfolio. The more I discover what I can do, the higher my expectations are of what kids can do. They seldom disappoint me. Second, so that your students have a quantity of writing and reading from which to make their selections, it's a good idea to ask them to produce, say, five rough draft pages of writing a week and to read for at least a half hour each night. Third, each trimester (other teachers might choose other time periods), ask your students to arrange their writing from most effective to least effective and to evaluate it by considering the following questions:

- What makes this your best piece?
- How did you go about writing it?
- What problems did you encounter?
- How did you solve them?
- What makes your most effective piece different from your least effective piece?
- What goals did you set for yourself?
- How well did you accomplish them?
- What are your goals for the next 12 weeks?

Finally, if you have to grade your students, base the grade on goals set

Adapting Portfolios for Large-Scale Use

Jay Simmons

Current large-scale methods of assessing writing ability operate with little real engagement by teachers or students. They also yield unnaturally low scores, especially for weaker students, fail to describe how writers of differing ability actually behave, and offer no instructional strategies tailored to specific achievement groups.

My research team in Durham, New Hampshire, has successfully pilot-tested an alternative to holistically-scored, timed writing samples that corrects these shortcomings without added cost. Our results indicate that self-selected portfolios of their best work are significantly better than timed tests in estimating students' writing abilities.

Our research team used portfolios to measure production, perception, and reflection (as suggested by Howard Gardner [Brandt 1988])[1] across a population of 27 randomly selected 5th graders. They chose their three best pieces finished during the school year and then wrote a timed-test sample, all of which we rated holistically.

For the portfolio assessment, we recorded such factors as paper length, mode of discourse, and range of modes included in the portfolio. Both students and raters listed strengths of the papers, so that we might see how well student reflections on the qualities of their papers matched those of adult raters.

When we compared the timed-test scores with the median score in the portfolio assessment, we found that timed tests significantly underestimated writing ability, especially of the weaker writers. Students who scored 2 or 3 on the timed test averaged 4.7 on the median portfolio piece, the same as those who scored 4 or 5 on the timed test!

Portfolio assessment identified writing habits associated with certain performance levels: those scoring highest produced longer, more varied portfolios. When students limited their range of modes of discourse, they relied on narrative, a form learned earlier in school. The lower the test score, the more the writer chose narrative exclusively for the portfolio.

Teachers can use the information that the portfolio assessment makes available about the behaviors and attitudes of writers of differing abilities to tailor their instruction. For example, those scoring lowest on the test seem to need vocabulary development and practice choosing shorter tasks when time is limited. The average writers in our 5th grade sample need to expand the range and depth of their writing choices, while emphasizing the overall flavor of the work. In narratives this may mean developing characters and motivating plot developments.

In 1989–90, Seacoast Educational Services of Somerworth, New Hampshire, a consortium of 12 school districts, will collect and analyze portfolios of 5th, 8th and 11th graders by randomly selecting a representative subset of the approximately 4,000 students in the population. Although both timed-test and writing folder pieces will be included, only one testing session will interrupt the instructional year.

Our field test scored just over 100 papers; the Seacoast project will handle more than 1,500. This will give the 12 districts more accurate and valid scores and more descriptive information about their students' writing. It will give their teachers statistical profiles of the writing behavior of the students and a more refined version of a naturalistic model for large-scale writing assessment through portfolios.

[1] R. Brandt, (December 1987/January 1988), "On Assessment in the Arts: A Conversation with Howard Gardner," *Educational Leadership*, 45: 30–34.

Jay Simmons is an English teacher, Oyster River High School, Coe Drive, Durham, NH 03824.

and achieved as evidenced in the portfolio of only their best work.

The Kind of Evaluation That Matters

Nahanni is not the exception in my class. She is becoming the norm. She is motivated and persistent, and she cares about learning. She reads and writes for real reasons. Through her portfolio, I know her as a reader, writer, thinker, and human being, not as a "5" or a "7." Through their portfolios, I know all my students. They, like Nahanni, are articulate learners because they continually *practice* discussing what they know and how they know it: with me, their peers, and with the community. They do so, for example, by sharing with other grade levels, teaching teachers at writing workshops, and publishing locally and nationally.

Learning to make meaning in writing and reading is not objective, as our evaluation systems would seem to indicate. We must become more flexible in our assessment of students' work. When kids are given choices in what they read and what they write, and time to think about what they are doing, their writing and reading get better. When we trust them to set goals and to evaluate their learning in progress, we will begin to realize that they know much more than we allow them to tell us through our set curriculums and standardized tests. If our goals are to keep students reading and writing, to help them get better at both, and to help them become independent learners, then we must nurture self-evaluation of writing and reading in progress. This is the kind of evaluation that matters—because it is for Nahanni, for Joel, for Sarah. Who else is evaluation for?□

Portfolios give teachers insights into their students' growth not possible through traditional assessment measures.

Reference

Costa, A.L. (April 1989). "Re-assessing Assessment." *Educational Leadership* 46: 2.

Linda Rief teaches 7th and 8th grade at Oyster River Middle School in Durham, New Hampshire. She is also an instructor in the University of New Hampshire's Summer Reading and Writing Program and has worked extensively with Donald Graves. She may be reached at RFD 2, Edgerly-Garrison Rd., Durham, NH 03824.

Reprinted with permission, from The American School Board Journal, April. Copyright 1991, The National School Boards Association. All Rights Reserved.

Show What You Know

The senior project requires students to prove their skills and knowledge

BY DEANA R. CHADWELL

PICTURE A TYPICAL high school on a late May day. Imagine the halls full of seniors, dressed like professionals, nervously encouraging one another. Picture a hundred (plus) community members—not parents, but local experts on anything from photography to child psychology—searching the hallways for their assigned rooms.

What's going on? It's the first day of senior project presentations at South Medford High School in Medford, Ore., and our seniors are facing the final hurdle they must clear before they graduate. After five months of work, they will demonstrate to community specialists their newfound expertise on everything from homelessness to investment banking. And they won't be alone: Following our lead, seniors in 60 schools nationwide will face similar hurdles this year.

A cure for senioritis

Some background: Five years ago, my colleagues and I were given the responsibility to develop a curriculum for a new high school. This was every teacher's dream. But then we thought about the seniors, those placid, horizontal creatures who maintain a hazy attention until the homecoming queen is crowned or MacBeth dies, whichever comes first, at which point they pass into the oblivion known as senioritis. We diagnosed their comatose state as evidence of an inadequate curriculum. After 12 years of dittoed handouts, they could hardly be thrilled at the prospect of more.

We also recognized that high schools all over the country were giving diplomas to marginally literate, ill-prepared seniors. So we asked ourselves, What should a graduating senior be able to do? Read? Surely. Write? Of course. Find information? Yes, this is the information age. Apply that information? Naturally—because information is of little use if people can't apply it. Speak? Definitely. And shouldn't high school graduates be capable of critical thinking, of independent learning, of intelligently managing their time? So why not require them to prove they can do all those things? Hence the senior project.

The senior project requires that all seniors, whether they're in remedial or honors classes, complete a three-phase program in their English classes before they are allowed to graduate. In phase one, they prepare and write a 10-page research paper on a topic of their choice. In phase two (call it practical application), they plan and carry out a project that relates to the research in some fairly obvious fashion, takes at least 15 hours of out-of-class time to complete, and is something they've never done before. In phase three, each student appears before a board of judges, more than half of whom are community members and all of whom have some expertise in the area the student chose to study. The students speak about what they learned from the project and field any questions the judges might ask.

So much for the theory. What do the projects actually look like? As you'd expect, they vary greatly, but all are ambitious. One student, for example, built a wooden hydroplane; another studied the ballets of George Balanchine and then choreographed a ballet of her own. Another student studied the problems facing Medford's homeless people, then staged a local version of a LiveAid concert to raise funds for an organization that works with the homeless. Still others refinished Victorian antiques, investigated new ways to invest scholarship funds, and reported on efforts to tutor learning disabled students. One senior even produced a time management booklet for other high school students.

Whatever the subject, the students are articulate and impassioned. After all, they've spent five months on their projects, and five months is a long time in an 18-year-old's life.

Brita Jackson, a South Medford High senior, rehearses a dance she choreographed as part of her senior project.

Deana R. Chadwell teaches English at South Medford High School in Medford, Ore. She and three other South Medford teachers are the principals in Far West Edge, Inc., a consulting group that provides workshops and materials for senior project programs.

BEST COPY AVAILABLE

The judges can't tell which kids come from remedial classes and which are honors students, because all have proved they can excel at something and all are proud.

Unexpected benefits

Not only has the senior project had a positive effect on our students, it also has done wonderful things for the school in general. It has helped us integrate the various elements of our curriculum, and it has done so naturally and effortlessly. Students write their research papers in English class, as in the past. But because we English teachers no longer control the topics, we often find ourselves asking our technical arts or computer science colleagues to read the papers for content. Teachers from all subject areas work together and serve as mentors for students in our individual areas of expertise, even if it isn't an area we necessarily practice at school. And all staff members—both certificated and classified—have the opportunity to sit on senior boards to serve as resident experts.

At South Medford High, too, we find that the senior project has had a trickle-down effect on the entire curriculum. If we were going to expect kids to speak confidently and articulately as seniors, we knew we had to beef up our speech instruction. When we discovered that students were having trouble citing sources in their papers, we strengthened those areas.

The senior project also has made more routine studies seem relevant to students. I can say to my freshmen, "Of course you need to learn to write a 500-word paper now. When you're seniors, you'll have to write a 3,000-word paper to graduate. So let's get started."

And what about the students who don't (or won't) complete the project? If schools increase and upgrade graduation requirements, some students (at South Medford High, it's usually only three or four each year) can't or, more accurately, won't meet those requirements. At least not right away. Occasionally a student will choose to pick up his diploma at the community college or take the exam for a graduate equivalency diploma. Each year, though, we have a handful of seniors who didn't graduate the previous spring but who return determined to complete their credits and their projects.

Putting things together

A program like ours is bound to have some problems, of course. We spoke to one school, for example, where the teachers were so excited about the project that all 50 of them rushed back to their classrooms and announced 50 different versions of what sounded like impending doom to shocked and horrified seniors. That night several hundred different versions of the calamity were related to parents, who in turn descended on the school board while the press watched. That school still hasn't adopted the program and probably never will.

Or consider the principal who has to tell the student government president that he won't be graduating because his research paper was plagiarized. Or the department head who has to work with the military to help a student who was twice foolish: once when he refused to do the project and again when he told the Marine Corps that he had a high school diploma. Or consider the English teacher who has to turn back dozens of rough drafts with "F's" on them. Holding the line isn't easy, but it's worth it. As one of our school board members put it: "The senior project gives solid evidence that our kids can meet society's expectations."

Starting senior projects

What's involved in starting a senior project in your own school system? According to South Medford High School English teacher Deana Chadwell and her colleagues, launching a senior project program requires a little money, a good bit of time, and no small amount of support from the local school board.

"The buy-in has to be complete," says Jay Monier, one of the teachers who developed South Medford High School's program. The senior project, he says, "simply won't work if there are discrepancies in philosophy" among teachers, administrators, and board members.

Once you have that support, though, the problems you're likely to encounter will largely be logistical. Chadwell and her colleagues report, for example, that community members typically are eager to take part in the senior project program, once they know about it. The big challenge is coordinating resources to make sure you have the right people in the right place at the right time.

One way to do this, notes Monier, is to designate someone—probably a teacher—as senior project coordinator with responsibility for overseeing the details and making sure, for example, that each youngster has a mentor for the project. The project director, in turn, should select an adult in the community to serve as community coordinator and liaison. (In South Medford, two former school board members serve in this position.)

Another tip: Provide a series of safety nets to help the youngster who might fail. "You need to know in advance," says Monier, "what you're going to do if the kid fails his research paper or if he's sick on the day of the senior boards." Some fail-safe provisions: Give opportunities to rewrite a failing paper or to schedule a second appearance before a senior board.

From a community standpoint, say Chadwell and others, the senior boards are a gold mine: What better way to let your schools strut their stuff? Parents, though, might balk. Explains Monier: "A senior's diploma depends on passing this project, and that fact just ignites some parents, especially those concerned that seniors have enough to do, what with the prom, after-school jobs, and preparing college applications." Against such criticism, perhaps the best defense is pointing out the academic benefits of the program.—*Donna Harrington-Lueker*

No one I know has any doubt about the need for change in the current structure of education. Recent research calls for more cooperative efforts among students, staff members, parents, and community—and says we must train students for lifelong learning. The senior project does all that and more. Indeed, the senior project promotes cooperative efforts on all fronts. It requires seniors to begin making their own decisions, and it treats all students—from remedial to advanced — as capable of achieving something great on their own.

RON BRANDT

Reprinted with permission from Educational Leadership, December 1987/ January 1988.

On Assessment in the Arts: A Conversation with Howard Gardner

The author of *Frames of Mind: A Theory of Multiple Intelligences*, Howard Gardner has spent years investigating issues of creativity, intelligence, and the artistic process. Here he explains the endeavors of Project Zero to develop assessment techniques for the arts. He contends that "making art" is central to artistic learning and that perception and reflection activities must be linked directly to student production of art.

You and your colleagues at Harvard Project Zero are working with Educational Testing Service and the Pittsburgh schools to develop ways to assess student achievement in the arts. Why?

We want to know whether individuals involved in the arts are getting something out of their experiences, and whether those results can be documented. I see school as a place to develop the different components of the mind. Artistic thinking—thinking in artistic symbols—is a distinctive way of using the mind, but one usually downplayed in school. For example, we use the mind in a certain way in creating or listening to music. You could work with mathematics or language for the rest of your life, and it would not affect your ability to deal with music. By the same token, the abilities involved in dealing with the visual arts—with sculpture or painting, with dance, mime, use of the body, and so on—all represent separate sets of cognitive skills. If we omit those areas from the curriculum, we are in effect shortchanging the mind.

Photographs by Constance Pfaff

Now, it might not be necessary to nourish these abilities in school if they were developed fully outside of school, but it's obvious that in most cultures—and certainly in ours—there's very little attention to the arts outside of school.

Is it really possible to assess achievement in the arts in a meaningful way?

Even people who are strong advocates of arts in the schools are very much in conflict about assessment. To some extent, they feel that what's most important in the arts is difficult to assess: arts have personal meaning, they deal with emotional content, they're often very subjective. So some people, including people involved in the arts, feel that assessment is incompatible with their purposes. Others, though, feel that we must assess results if we are to justify the expense of offering these experiences in schools. But they often make the opposite error of trying to assess the arts in the same way other kinds of learning are assessed: with multiple-choice tests and so on.

The assessment techniques you're developing are for use with all students, not just those with special talents?

Exactly.

Are they intended as measures of ability or achievement?

In the beginning, we had the mistaken idea that one could assess aptitude apart from intensive involvement with material. We're convinced now that that probably doesn't work anywhere, but it certainly doesn't work in the arts. It's pointless to test people's musical or dramatic aptitude unless they've had a significant amount of experience in artistic thinking. And many kids—even at the high school level—have had virtually no exposure at all. So our approach has been to involve students in very rich kinds of artistic activities. We involve them not only in artistic creation—actually painting, composing, and so on—but also in what we call "projects." These are curriculum modules that involve fairly long-term kinds of analysis and perceptual discrimination, as well as considerable "reflection" on what they're doing. And after students have had some experience with these kinds of things, we assess them, so to speak, "on the fly" while they're actually engaged in these projects.

As an analogy, let's say that somebody wanted a chess aptitude test. We could give people a chessboard and say, "All right, play with these for 20 minutes, and we'll see how good you are." Or recognizing that they don't know the rules of chess, instead we could see how good they are at counting, calculating, spatial reasoning, and so on. But you would probably find that such related abilities are a poor predictor of actual chess playing, because chess is a complicated activity that also involves "psyching out" what's on your opponent's mind, thinking several steps ahead, that sort of thing. So it's much more sensible to take a hundred people who have never played chess before, teach them the rules, have them play for a while, and then begin to give them chess problems.

That's basically what we're doing: we're assuming that most kids have not had much experience artistically. We immerse them in activities that call for artistic thinking, and after they have some familiarity with an activity, we use various kinds of assessment techniques, including some straightforward ones as well as some that are more unobtrusive.

How can you be sure about the validity of these new techniques?

Well, to begin with, they're not all new. Both Project Zero and ETS have had extensive experience that tells us what sorts of things are *likely* to work and not to work.

But a crucial part of this effort is that, from the very first, we have worked directly with teachers. In workshops we conducted in Pittsburgh, the teachers became familiar with our materials and began to do exercises on their own, and then slowly began to work with students while we looked over their shoulders, trying to make sense of what the students were doing.

"If we omit [the arts] from the curriculum, we are in effect shortchanging the mind."

The first step, in other words, has been to find out whether the materials we developed were acceptable to teachers themselves, which is often not a trivial matter because some of these ideas are not very transparent. Also, the teachers often make useful suggestions and modifications. Once the teachers think the exercises are usable, they try them with students. Then, if they work, we find that the teachers are usually able to reach general agreement on how to analyze the results.

A good example is music. Because we believe it's important to work with teachers, we started with what music teachers were concerned with: performance. We came up with a set of measures to look at changes in student performance over time: how a clarinet player or an ensemble improves. I am pleased to say that in our first go-around with teachers there was actually fairly high agreement on how to score the results.

Similarly, in the literary area we are developing a complex set of instruments to monitor a variety of different aspects of student writing. We began with poems. We gave the kids a poetic frame with which to work, and then monitored their figurative language, their sense of voice, and so on—there are a lot of things to look at. Again we've been pleased with the initial responses. The teachers have not only caught on to the system but are excited by it.

These measures are intended for use by teachers, then, not by external evaluators?

Our goal is to come up with instruments that make sense to teachers in regular schools and that will provide useful feedback on how students are doing. That's been the goal of the Rockefeller project in general. We're not part of the current critical thinking initiative, but this effort is analogous to it.

Whether there will eventually be a set of National Assessment-type instruments that could be administered without intensive participation by local teachers is a question that hasn't been answered yet. Clearly, that is the sort of thing that ETS might want to do, and I'd be glad to see it happen if it seems feasible.

What are some of the factors to consider in assessing aesthetic growth?

We think one should monitor three kinds of things: production, perception, and reflection.

This idea isn't highly revolutionary, but it's more unusual than one might think. This past spring I spent three months in China. We have a great deal to learn from the Chinese about teaching students to produce great works in the traditional way. But they haven't confronted at all the issue of how people learn to *see* art nor of how they reflect on it.

Now, there's a growing group of American art educators who say our schools spend too much time on producing art and that we ought to teach the arts as we do other disciplines: take a verbal and analytic approach—emphasize historiography, aesthetics. I don't think that's necessarily a bad thing, but there's a great danger that it will slip into the traditional ways of knowing, verbal-analytic and logical-mathematical forms, which are covered pretty well in the rest of the curriculum.

From the very beginning, Gardner explains, teachers have contributed their ideas and judgments about assessing student work in the arts to Project Zero.

Our approach differs from both the Chinese and the American in two ways. One is that, particularly with the younger kids, production continues to be central. That is, we think artistic learning should grow from kids *doing* things: not just imitating, but actually drawing, dancing, performing, singing on their own. And I mean not just songs they've been taught, but singing their own compositions. That's central to our approach—and it's very different from just learning traditions from the past or just talking about art.

The second aspect of our approach is that, as much as possible (and our exercises strive to do this), production should be linked intrinsically to perception and reflection. Perception means learning to see better, to hear better, to make finer discriminations, to see connections between things. Reflection means to be able to step back from both your production and your perceptions, and say, "What am I doing? Why am I doing it? What am I learning? What am I trying to achieve? Am I being successful? How can I revise my performance in a desirable way?"

Now, that's quite different from the "discipline-based" approach, which always proceeds in terms of the standard disciplines. Our approach grows out of the child's actual experience with the arts. We don't talk about perception or reflection apart from artistic *activity*. As kids grow older, those aspects can be taught more intensely. I have no objection to teaching art history to 10- or 12-year-olds, but I do object to teaching it apart from actual artistic production.

An example we use describes a student at work on a triptych: portraying images of three faces in a tripartite work. After working on the project for several weeks, she was introduced to the way artists in the Middle Ages put together triptychs. That way the student was much more interested, because she herself had been trying to solve the same problem of producing three distinctive yet complementary forms.

Another example, which I used a lot in China, is theme and variations. You take something like "Happy Birthday" or "Twinkle, Twinkle," which everybody knows, and you make sure kids can sing it or play it on a kazoo. Then you have them do "Twinkle, Twinkle" in different ways: do it sad, do it happy, do it spring, do it winter, do it limping, do it in 3/4 time, and so on. Then you show them that, in fact, Mozart also wrote variations on "Twinkle, Twinkle, Little Star." Again, the children are interested because they've been confronting the same problem as Mozart.

That type of exercise would seem to be generalizable across disciplines.

Many of them are, but not all. Exercises in pitch or in color arrangement don't have direct analogies in other art forms. But at Project Zero we've been working simultaneously or in parallel with a number of different art forms. We try to get maximum applicability but, as you may know, my theory of multiple intelligences holds that no two forms of thinking are exactly comparable.

I'd like to ask about your theory. First, though, will you give one or two more examples of ways to assess growth in the arts?

A good example that begins with perception is the Mona Lisa exercise. We show some slides of fakes of the Mona Lisa along with the original, and we get the kids very interested in which is which. We show a slide and say, "What is this?" . . . "It's the Mona Lisa." . . . "Are you sure?" Some kids say one thing and some another. They talk about the differences they see and the reasons. After that kind of reflection, we get them experimenting on their own, trying to either copy a work like the Mona Lisa or copy things like signatures, so they begin to see what is involved in doing something so that it looks as if it was done by someone else. Now, that task begins with perception, passes through reflection, then eventually goes to production. And we go from a fun puzzle to the serious question of style.

"We think artistic learning should grow from kids doing things: not just imitating, but actually drawing, dancing, performing, singing on their own."

Another form of assessment involves the building of a portfolio. Over a period of time, students accumulate their various efforts, including drafts, notes, false starts, things they like and don't like. Their portfolio becomes a kind of data base that both teacher and student can look at to see what's been done and what's been learned. Some of the measures can be quite explicit: how many entries kids make in their notebooks, how detailed they are, things like that. But others can be more difficult to get exact reliability on; for example, how many different sources of information have been used in trying to improve a work, or when a work has deepened, rather than simply changed.

We think these kinds of data should be enlightening to students, even if the teacher never looks at them. But of course the teacher does look at them too.

I see a lot of implications in this project for teaching and for curriculum design.

And not just in the arts. Almost everybody realizes that the American schools have been disappointing in recent years. But I think most of the reactions to this concern will not be very productive in the long term. Getting higher scores on standardized tests is not the real need, and at best it responds to symptoms rather than to the underlying disease. What we need in America is for students to get more deeply interested in things, more involved in them, more engaged in wanting to know; to have projects they can get excited about and work on over longer periods of time; to be stimulated to find things out on their own. In a way, the arts are a good testing ground for such activities because many members of the educational establishment don't care about them so much, so teachers can afford to take more chances.

A little earlier, you mentioned your theory of multiple intelligences. How does this arts assessment project relate to it?

Well, nobody has to buy my theory to buy the project, and in fact it's only tangential. But it is related in the sense that program officers at the Rockefeller Foundation, our sponsors, felt that many young people are being cheated out of the opportunity to receive a good education—by college admissions criteria and so on—because, while they have many abilities and many achievements, they tend not to do well on standardized tests. There's a lot of agreement about that, even among the testing experts. Having a certain kind of intelligence we might call "test smarts" helps some people scoot through school very well, but everybody knows that it doesn't have nearly so much predictive value for what happens outside, or after, school.

So the Rockefeller Foundation became interested in the possibility that school has focused too much on a certain combination of intelligences—linguistic, logical, mathematical; but that life certainly involves more than that, so school should be concerned with more than that too. They wanted to see whether we could find ways to pick out the kids with skills in some of these other areas, so those kids would have a better chance to get into a good college and to achieve success in life, and would also acquire a better self-concept in the process.

For those who may not know your theory, would you summarize it briefly?

Evidence from a variety of sources, especially new information about development of the nervous system and organization of the brain, indicates that human beings have evolved over a long period of time to think in at least seven ways, which I call "intelligences." I mentioned two of them—mathematical and linguistic—but there's also musical, spatial, bodily-kinesthetic, interpersonal, and intrapersonal intelligence.

Everybody who's normal has the potential to develop each of these intelligences to a large extent, but we don't all have the same profile to begin with, and we certainly don't all end up with the same profile. Most important, strength in one intelligence does not have predictive value for strength in another intelligence. You can be very strong in music and not in spatial, or vice versa. Therefore, psychologists who claim that intelligence is a single entity—educators who think of students as either smart or dumb—are wrong. A person can be smart in one area and dumb in others.

Part of our effort is trying to find how best to define each of the areas and how best to assess aptitude and achievement in them. We're also interested in which forms of intelligence are important in the various arts. For example, linguistic intelligence plays a role in the study of science and history and geography, but it's of central importance in writing fiction or poetry.

As a psychologist, what made you decide to define intelligence this new way?

It was probably less as a psychologist than as a participant in American public school life that I felt intelligence had been appropriated to refer to a very narrow band of abilities. And as long as the word is saved for just those abilities, people have the license to call everybody who doesn't have those abilities "stupid." So I deliberately tried to use the word in a new way to make people realize that there were a lot of tacit assumptions that needed to be confronted.

"Getting higher scores on standardized tests is not the real need . . . what we need in America is for students to get more deeply interested in things, more involved in them, more engaged in wanting to know . . ."

I would have no problem with calling intelligences "talents," as long as people also called language and logic "talents." But they don't; they say that's being smart. The fact that I've annoyed many people by using the word this way is a kind of perverse confirmation that people thought they owned the word and that it had to be used their way.

How is the theory being accepted?

Well, the closer people are to psychometrics, the more they don't like the theory at all—and they probably have good and bad reasons for that. It's a threat to a lot of people who have an intellectual and a financial stake in a certain way of looking at things. I'm pleased by the fact that people who I think are good scientists but who don't have a vested interest in a particular orientation often like the theory. Biologists think it's sensible. And even psychologists—those who aren't wedded to psychometrics—feel it's a useful point of view. Many educators find it appealing. It's not a theory with a capital T; it's a new way of thinking about things, a new way of organizing a lot of information.

A skeptic might say that it is just another way of describing what we've always known: that people differ in their abilities.

I think it's taking two forms of intelligence off their pedestal.

The skeptic might also say that, far from ignoring the other forms of intelligence, schools do a lot to foster them. They certainly pay attention to athletic talent, for example. They do teach music and art. They encourage kids in their social activities, and so forth.

I can't go along with that. Music and art are the first to go when there's a financial crunch. As kids grow older, they get less and less of it. It may be important in kindergarten, but it disappears by eighth grade, except for the kids who are professionally oriented. Probably the most important reason I disagree is that even though there's a lot of focus on the kids who are good performers—who play an instrument well, for example—teachers don't often engage them in musical thinking, in visual thinking. I can only say that these are new ideas to many of the art teachers I work with.

If your theory were more widely accepted, how would that affect school practices?

I think that it is beginning to be accepted. Education is gradually becoming more individual-centered. Surely we've passed the point where we want to teach every kid the same thing the same way. My theory gives educators a way of thinking about individual gifts and how to accommodate teaching to them, but also a way of teaching conventional subject matter in ways more accommodating to the different ways of knowing. A student's profile of intelligences ought to have some effect on the way he or she goes about learning history. Another implication is the realization that, while assessment is important, we need to be much broader in what we assess and much more flexible in how we assess it.□

Howard Gardner is Co-Director, Project Zero, Harvard University, Graduate School of Education, Longfellow Hall, Cambridge, MA 02138. **Ron Brandt** is ASCD's Executive Editor.

Reprinted with permission from The Reading Teacher, Vol. 44, No. 9, May 1991.

ASSESSMENT

Portfolio assessment for young readers

Sheila W. Valencia, University of Washington
Guest author: Scott G. Paris, University of Michigan

The overarching goal of new approaches to literacy assessment is to change the ways that teachers and students interact in the classroom. Ideally, new assessments will increase the engagement of students in thoughtful reading and writing activities and will provide more detailed records of students' literacy development over time. Portfolios should also inform teachers about the interactive dimensions of literacy and make them sensitive to processes of learning rather than just the outcomes. Portfolio assessments empower teachers to select and create measures of students' talents and weaknesses so that appropriate instructional opportunities can be provided for different students. Thus, teachers are often enthusiastic but bewildered by the challenge of creating their own portfolio assessments. In this brief essay, I would like to highlight some of the promising avenues to pursue as teachers strive to create alternative literacy assessments.

Valencia (1990) points out that effective portfolios of literacy assessment involve authentic activities that promote collaborative reflection. Developing abilities should also be measured repeatedly with a multidimensional variety of tasks. In this manner, portfolios assess cognitive, motivational, and social processes that underlie literacy development. Assessment is embedded in the curriculum and interwoven with instruction so that strategies can be assessed and improved. Self-perceptions of competence and control can be encouraged, and appropriate individual mastery goals can be set and pursued.

During the past year, several students and I have worked with teachers and administrators to design portfolio assessments for reading in primary grades. We quickly discovered that most published tests were not consistent with the philosophy of the district, so we examined different research tasks used to measure comprehension strategies, metacognition, motivation, and literate environments. We shuffled through tasks around a large table and negotiated an agenda that none of us could have articulated beforehand, a process that was aptly described by Donmoyer (1990) as the "deliberative approach" to curriculum evaluation. After many hours of reviewing, creating, and revising tasks appropriate for young children, we created a battery of different tasks that were administered to children in six different schools. Our experiences and data revealed that some tasks were on target and appropriate while others were fundamentally flawed. We were pleased to find that several of our tasks provided robust measures of young children's strategic reading and metacognition. These seem particularly worthwhile to share because they can be adapted easily to portfolios at any grade level.

Strategic reading, comprehension, and metacognition. We used the Strategy Assessments in *Heath Reading* (1989) as a model for measuring children's understanding of strategies before, during, and after reading. These prototypes include examples of relevant questions to ask students as they read and thus provide a "think aloud" procedure for assessing comprehension and metacognition simultaneously. The procedure avoids asking young children hypothetical questions that often elicit fabricated or socially desirable answers.

We modified the tasks by asking five uniform kinds of questions: identification of the topic from text cues before reading, predictions of the content based on previous passage information, inferences from text, monitoring the meaning of a novel word, and summarizing key information. In addition, children were asked five follow-up questions such as "How did you know that?" or "If you did not know that, how could you find out?" Children's responses to these metacognitive questions were given one point if they explained such things as how they used the titles for identifying the topic, how prior text influenced their predictions, how rereading or skimming might help them make an inference, and how context provides clues about the meaning of an unknown word.

These ten questions, five comprehension and five metacognitive, were woven into a conversation as children read aloud the passages, one narrative and one informational. As children read the passages and answered the questions, we recorded the frequency of strategies observed (e.g., lookbacks, questions, elaborations, rereadings, etc.). Good readers generally exhibited three observable strategies as they read each passage. Children who scored at least 3 of the 5 points for comprehension questions and 3 of the 5 points for their metacognitive explanations were found to be using appropriate strategies and displaying good awareness of their own thinking.

The high correlations among the measures of comprehension, metacognition, and observed strategies indicate that children's abilities to explain how they did answer questions

BEST COPY AVAILABLE

are related to their comprehension. Thus, the task provides useful information about reading comprehension strategies and thinking that has been difficult to obtain for young readers. We also designed tasks suitable for kindergartners and gave the passages as listening rather than reading tasks to children who could not read adequately. The tasks worked equally well as reading or listening comprehension assessments.

There are several distinct advantages to the strategic reading task that make it useful as a prototype for portfolio assessments. First, the strong relations among strategy use, comprehension, and metacognition have been shown empirically. Second, the task can be presented as either a listening or reading task and is therefore suitable for children with different abilities in kindergarten and first grade. Third, the procedures can be modified so that the passages and strategies are taken directly from the classroom curriculum. This means that the assessment is not an added burden to the teacher but is embedded in regular instruction. Fourth, the task can provide cumulative evidence of children's abilities to use reading strategies that have been taught in the classroom.

Additional measures of comprehension. Two other tasks seem promising for measuring children's comprehension: cloze procedures and retelling. Cloze tasks are highly correlated with age and awareness of print conventions in our studies of young readers and, additionally, they are consistently correlated with measures of strategic reading in older children (Paris, Cross, & Lipson, 1984). Cloze tasks are particularly well-suited to comprehension assessment for several reasons: They can be designed by teachers with regular curricular materials; they can be made easy or difficult by varying the numbers of options provided for each missing word (in the slots or word banks); they can be given as reading or listening tasks; and they can be given to groups of students and scored quickly.

Retellings are much more labor intensive but can be used flexibly with curricular materials as reading or listening tasks so that they also become embedded in instruction. The *Qualitative Reading Inventory* (Leslie & Caldwell, 1990) provides grade-level passages and guidelines for assessing comprehension using oral retelling (as well as running records and comprehension questions).

Additional measures of metacognition. We also created a Book Selection task, based on Kemp (1990), in which children engage in thinking aloud as they identify books that they would like to read. We encouraged children to tell us how they decided if they liked the books. We recorded their comments and checked their use of overt strategies such as skimming the book, making affective statements, or talking about previously read books or prior knowledge about the subject. After children had sorted books into two piles, we chose one book from the preferred stack and used it to assess children's concepts about print.

Our interview was modeled after Kemp's (1990) procedures and Clay's (1979) *Sand* and *Stones* tasks. In our pilot study, the children's think-aloud comments and their concepts about print were related to age and comprehension. The virtue of this task is that it provides an assessment of children's conceptual orientation to text that is often not available for young readers. Teachers have flexibility in the materials and concepts they select so that the assessment is developmentally appropriate for children who may not read much.

Motivation. We created a motivation-for-reading questionnaire that included questions such as "Are you a good reader?" and "Do you enjoy reading?" Questions about young children's abilities, interests, and enjoyment were answered with optimism and enthusiasm by nearly all children and, thus, were not related to other aspects of reading behavior. The positive, nondiscriminating attitudes are consistent with developmental research on young children and may be due to optimism, poor self-evaluation, or naiveté. Many other attitude scales reveal similar skewed responses by young children (McKenna & Kear, 1990) which suggests that decontextualized attitude surveys may not provide useful data for young children. We urge teachers to use behavioral descriptions of motivation and attitudes such as effort, challenge seeking, and goal setting, in order to substantiate their judgments about students' motivation.

Literate environments. We designed several indirect measures of literacy backgrounds of the children. For example, we created two lists of well known children's books and authors (that also included nonbooks and nonauthors) and then asked children to circle any of the books and authors that they had heard about before. This simple recognition measure (based on research by Stanovich & Cunningham, in press) was highly correlated with children's ages and with comprehension and concepts about print tasks. Our survey of children's literacy habits beyond school was not highly correlated with other measures of reading. Although literacy practices at home may be related to children's reading and writing, surveys of young children may not provide sensitive measures of the quality of their literate environments.

In summary, we believe that portfolios of literacy development offer exciting opportunities for teachers to create assessments that engage students in authentic activities and provide genuine guidance for instruction. Some of our tasks can be adapted easily by teachers to fit their needs and curricula. Here are some simple suggestions: Start slowly. Try two or three new tasks with another teacher. Develop a system for keeping records and sharing the information with students and parents. Use portfolios to stretch your teaching into new areas of literacy, for example, strategic reading, metacognition, or motivation. The new activities may open new horizons for students and teachers alike.

References

Clay, M.M. (1979). *The early detection of reading difficulties.* Auckland, NZ: Heinemann.

Donmoyer, R. (1990). Curriculum evaluation and the negotiation of meaning. *Language Arts, 67,* 74-86.

Health Reading. (1989). Lexington, MA: Heath.

Kemp, M. (1990). *Watching children read and write.* Portsmouth, NH: Heinemann.

Leslie, L., & Caldwell, J. (1990). *Qualitative reading inventory.* Glenview, IL: Scott Foresman.

Reprinted with permission from Phi Delta Kappan, May 1989.

Assessing Science Learning In Elementary School: Why, What, and How?

By focusing more clearly . . . on the outcomes we wish to achieve, we can effectively change the quality and quantity of science instruction in elementary schools, these authors suggest. That's because teaching and assessment go hand in hand.

BY SENTA A. RAIZEN AND JOYCE S. KASER

HERE ARE some items that might be used to assess a student's knowledge of science. Although they are fairly typical questions for gauging a student's learning in science, they are very limited in what they tell us about a student's knowledge of and skills in applying science.

Item 1 [illustration of four animals, one of which is a bird, while the other three are mammals]
Which one of these animals is not a mammal?
a.
b.
c.
d.

Item 2
Which of the following is used to measure temperature?
a. feet
b. degrees
c. centimeters
d. minutes
e. calories

Item 3
Mary and Jane each bought the same kind of rubber ball. Mary said, "My ball bounces better than yours." Jane replied, "I'd like to see you prove that." What should Mary do?
a. Drop both balls from the same height and notice which bounces higher.
b. Throw both balls against a wall and see how far each ball bounces off the wall.
c. Drop the two balls from different heights and notice which bounces higher.
d. Throw the balls down against the floor and see how high they bounce.
e. Feel the balls to see which is the harder.

SENTA A. RAIZEN is director of the National Center for Improving Science Education, Washington, D.C. JOYCE S. KASER (George Washington University Chapter) is director of the Washington, D.C., office of the NETWORK, Inc., which has its headquarters in Andover, Mass.

Illustration by Jean K. Stephens

Each of these three items could be found on either a standardized or a teacher-made test, the two types of testing common in elementary schools across the country. A brief review of the purpose and design of standardized instruments and of the tests and quizzes prepared by teachers will give readers some insight into the strengths and weaknesses of assessing science learning today.

CURRENT STATUS OF ASSESSMENT

Standardized tests in elementary science are now given in more than half of the states, usually in grade 3 or grade 4.[1] The test to be used may be purchased from a commercial source or it may be developed by science educators specifically for the state. Usually, test results are interpreted in one of three ways, all of which involve a comparison of some sort.

First, test results for a state (or for a district) might be compared to some absolute standard. For example, a group of educators may have determined a level of science knowledge and competence that elementary students should have acquired by a certain grade. These prespecified achievement criteria can be used as the reference standard for interpreting test scores. One difficulty with such criterion-referenced tests is that establishing the performance criteria is an arbitrary process and often leads to setting minimum standards, which could attenuate further an already weak science curriculum.

Second, test scores for current students can be compared with those of students at some time in the past. For example, a state or district might ask how the scores of fourth-graders who took the science test in 1988 compare to the scores of students who took a comparable test in 1984. A number of states are establishing trend lines of this sort for their students' science learning, as the National Assessment of Educational Progress (NAEP) does for the nation.[2] A difficulty with this approach to interpreting test results is that preserving comparability between tests sometimes makes it impossible to match the assessment to curriculum innovations that may have been introduced in the meantime.

Third, test data can also be used to compare science performance among schools, as in California; among different population groups or regions of the country, as the NAEP has traditionally done; among states, as the NAEP will do in the future; or even among countries.[3] One problem that arises, of course, is making these comparisons "fair," that is, taking into account various background variables known to affect student achievement in science.

Policy makers are interested in data from standardized tests as indicators of student learning in science, but, no matter what sort of comparison is used, the data are limited. First, because these tests are generally timed paper-and-pencil instruments using a multiple-choice format, they focus on facts and, to a much lesser extent, on certain types of deductive reasoning skills. Certainly, factual information is important. Indeed, one cannot "do science" or engage in scientific reasoning and problem solving without having a solid base of factual knowledge. However, students may be able to do well on many standardized tests by memorizing the requisite facts of science without being able to apply them in any way. Moreover, the tests are not informative regarding students' ability to conduct scientific investigations. Being able to identify or describe a scientific procedure – observing, measuring, recording data – on paper does not mean that an individual knows when that procedure is appropriate or knows how to carry it out.

Second, the range of knowledge of science that is tested may not match a district's or a state's curriculum. This is particularly true of commercially developed tests. Since these tests are administered to thousands of students across the country, the items generally are designed to assess core curriculum – the information that the largest number of students will have been taught. Trying to capture this core often leads developers to produce a test of unrelated facts covering many diverse topics. If students have had science instruction in depth in a few key topics – a teaching strategy much more likely than rote memorization to lead to real understanding – they are likely to be penalized by the usual sort of test, despite their deeper knowledge of science.

Finally, these tests provide a "snapshot" of a single point in time; they do not demonstrate growth in science knowledge and performance. Through allowing only one right (and preselected) answer, they mirror a conception of science knowledge and understanding that is fundamentally in conflict with the canons of science, which hold that scientific knowledge is always tentative and open to change based on new evidence.

Even though these limitations of standardized tests are widely known, test score data remain seductive: Richard Murnane and Senta Raizen have noted that, "when data are disaggregated on any basis and presented side by side on a page, the temptation to make evaluative comparisons, whether warranted or not, is overwhelming and nearly universally succumbed to."[4] This makes it critically important to work on improving standardized tests, so that they do a better job of reflecting all the important goals of science education.

Given the limitations of such tests, perhaps it is fortunate that classroom teachers seldom use the results of externally imposed standardized tests for their own purposes.[5] Instead, teachers conduct their own assessments of student learning, usually in order to assign grades. Teacher-designed tests can serve many other purposes, however. They can be used to determine what children bring to a science lesson, how much of the instruction has been effective and with which children, what impediments to learning need to be overcome, and what content a child does not yet understand. Using assessment in this manner, a teacher will have some guidance in how to modify instruction.

Unfortunately, assessment in the classroom seldom serves these important instructional purposes. Although they have a wide repertoire of approaches to assessment to draw on, including keeping records of and observing children's performance in science, teachers usually model their own tests on the end-of-chapter quizzes in textbooks or on the items in standardized tests. When teachers assign grades, they may or may not supplement such testing by considering the children's homework and lab work – if any lab work was done.

Items such as those with which we began this article can be found on either standardized or teacher-made tests. Note that the first item asks for recall of facts only. A student could get the right answer simply by having memorized which familiar animals belong in which category. Whether the student understands the distinctions that give rise to such classification or knows what to look for in order to classify a strange animal remains unknown.

On the second item, which also asks for recall of factual information, students could get the right answer without ever having had any instruction in science. By drawing on the everyday experience of

having their temperatures taken when they are ill, students could easily determine that "b" is the only possible answer.

Although the third item is more complex than the first two, a teacher would find it difficult to determine why a student got the right or wrong answer. There is no way to assess the reasoning that went on in the student's mind in answering the question.

Benjamin Bloom sums up current testing practices as follows:

> [T]eacher-made tests (and standardized tests) are largely tests of remembered information. . . . [I]t is estimated that over ninety percent of test questions the U.S. public school students are now expected to answer deal with little more than information. Our instructional material, our classroom teaching methods, and our testing methods rarely rise above the lowest category of the [Bloom] taxonomy – knowledge.[6]

The fact that many elementary school teachers assess only superficially what their students have learned in science is to be expected. One of the reasons that the assessment of science learning in this country is disappointing is that the instruction itself is weak. Science is taught in a larger school context – a context that holds that the basic skills are most important. And science is not generally considered a basic skill. Teachers often spend so much time on reading, mathematics, and writing that science gets short shrift. For example, after conducting a nationwide survey, Iris Weiss reported that the average primary teacher spends 19 minutes a day on science; teachers of grades 4 through 6 devote an average of 38 minutes a day to science, with lecture and discussion taking up more than three-fourths of that time.[7]

The problem of the lack of time devoted to science in elementary schools is compounded by teachers who don't know very much about science and who lack the confidence to engage in authentic inquiry with their students. This situation is to be expected, given the poor preparation that teachers receive in science. Often a single three- or six-credit introductory college course (usually a biology course) is all the science that an elementary teacher takes, and there are few opportunities to receive inservice training in science.[8]

As a result, textbooks have become the science curriculum, and much science learning is passive and superficial. This kind of learning is reinforced by tests that assess primarily factual recall and rote problem solving; the tests rarely require a deep understanding of science. Nor can teachers be faulted for constructing tests that are limited. Skill in testing is hardly ever seen as an important competency for teachers to acquire.[9]

During the last five years, the improvement of science education in the nation's schools has attracted a great deal of interest. Policy makers are not pleased that U.S. students perform so poorly on national and international tests – limited though they may be. It appears that students in this country do not have as much knowledge of science as students in other developed countries.[10]

However, an objective observer might ask, Why should they? The amount of time spent on science in elementary schools is limited, teachers are not trained to teach science, and the usual forms of assessment do not address some of the most important goals of science education.

There are a number of ways to begin changing the poor performance of students in science, among them: improving teacher training, setting aside time specifically for science instruction, developing better textbooks, equipping laboratories better, and providing administrative support. But the *assessment* of science education is also a possibility, although assessment is not often viewed in this way. Yet assessment can be a powerful tool for reform, since changing the nature of assessment can lead to changing the nature of instruction. If assessment were to be used as a lever for improving science education, what changes would need to be made in today's assessment practices?

TOWARD BETTER ASSESSMENT

A more helpful and comprehensive approach to assessing students' knowledge, skills, and abilities in science would have to address both the improvement of tests designed for large-scale administration and the improvement of teacher-controlled assessments. We must also pay attention to the correspondence between tests intended for policy use and those intended for classroom assessment, so that they send the same messages about the goals and expectations for student learning.

Tests designed for large-scale administration. Even within the limitations of time, cost, and consistency of scoring, there is much that could be done to improve tests intended to provide information about large numbers of students. Multiple-choice items ought to concentrate on important knowledge and concepts and should probe deductive reasoning skills. Some open-ended questions should also be included to permit students to create their own responses. Experience in Great Britain and some experiments in this country have shown that reliable scoring protocols can be developed for well-designed open-ended items. Most important, some standardized performance tasks need to be included. Several states, including California, Connecticut, Illinois, and New York, are now mandating the inclusion of such tasks.[11] Apart from individual items, the test as a whole should represent an adequate set of probes that match the important goals of student learning and competence in science as set forth by a district, a state, or a national body of experts, such as the goals outlined by Project 2061 of the American Association for the Advancement of Science.[12]

The domains being tested can be expanded by using matrix-sampling techniques of test administration: that is, administering overlapping sets of test items to samples of students, with no student taking the entire set of items. This technique is appropriate for addressing questions about science achievement in a district, in a state, or in the nation, since complete individual scores are not necessary for this purpose. We cannot stress strongly enough the importance of con-

sidering what science knowledge and competencies a test actually assesses before using it as an indicator of science learning. What do test scores matter if what the test tests is not important?

Teacher-controlled assessments. Assessment in the service of classroom instruction would consist of 1) teacher-made tests, 2) tangible records of students' work, and 3) teacher observation and documentation of student performance.

1. It is certainly appropriate for teacher-made tests to contain some items that probe students' knowledge of facts and routine problem solving. However, such tests are not sufficient for assessing what students know about science. Tests should also contain items that require students to exhibit productive thinking skills and to apply their knowledge of science.

Teacher-made tests should probe the process by which students solve a problem, rather than simply record the answer. In fact, there may be more than one "right" answer for a problem, and certainly there will be more than one way to arrive at an appropriate answer. Teacher-made tests should provide some way to track these processes. In addition, students should have opportunities to demonstrate not only their core knowledge of science, but also their skill in using the tools of science in hands-on activities. This means that some performance tasks should be part of any assessment.

An effective assessment of science learning in the classroom should also give students the opportunity to sort through disorderly situations in which the problem to solve is not indicated and to discover their own questions or problems. Some of these situations should be presented and assessed as group activities.

Designing a better science assessment is not without some disadvantages. Constructing the kinds of items we suggest takes more time than constructing a true/false or multiple-choice quiz. Administering the test requires a greater investment of time, since the questions are more probing. And because the answers are more complex and the processes for arriving at them are part of the assessment, the scoring takes longer, as well. The information that the teacher gains, however, is so much more complete and useful that, in our view, it more than justifies the extra time. Besides, once the initial work has been done, perhaps by a team of teachers with the help of a science supervisor, such tests can be adapted readily by individual teachers to suit their classrooms.

In the accompanying sidebar is a list of questions to ask in evaluating a science test – either a teacher-made test or a standardized test that a district or state is reviewing for use.

2. A second way a teacher has of assessing students' knowledge of science is through tangible records. These can include records of homework assignments – not just whether the student completed the assignment but what the student did or did not know. A science project – a written report, a model, or a demonstration – can also provide evidence of achievement. Students' laboratory notes, journal entries, or records of experiments carried out (including a description of the process) are also potentially useful in assessing achievement in science. Including records of hands-on work in assessments highlights the importance of laboratory work and application.

The reason for the emphasis on hands-on activities and laboratory work is that the interplay between factual knowledge, understanding of scientific constructs, and practical application helps students understand what they are learning. Here is an example of a situation that can be used in a variety of ways for assessment or instruction.

The owner of a local restaurant has been receiving complaints about the quality of the paper towels provided. He asks a fifth-grade class to conduct some tests to determine the best brand of paper towels for him to buy. He provides his criteria – the towels must be strong, absorbent, and not too expensive.

To answer his very practical question, the students will have to design tests to see how different brands of towels perform on the various criteria, record their procedures and findings, report them, and then make a judgment on which brand of towel to recommend. In the process, they will have to draw on a wide range of science knowledge, science skills, and problem-solving abilities. As teachers work with students on such a project, they are in a good position to collect records of the children's work, observe their performance, and assess their knowledge and skills.

3. The third way that teachers can assess what students know about science is through systematic observation. Teachers do use observation; they are aware that Janine almost always has the right answer, that Jacob often disrupts the class with his comic antics, and that Christine sometimes pays attention and sometimes doesn't. However, the systematic use of observation for purposes of assessment is less frequent. Here are two examples of such observations.

• Sometimes setting up a problem is the most difficult part: weeding out extraneous information and figuring out what problem needs to be solved and what information is needed in order to

Questions to Ask About Tests

HERE ARE some questions that the National Center for Improving Science Education suggests that teachers and administrators ask when they evaluate the quality of a science test.*

1. Are there problems that require students to think about and analyze situations?
2. Does the test feature sets of problems that call for more than one step in arriving at a solution?
3. Are problems with more than one correct solution included?
4. Are there opportunities for students to use their own data and create their own problems?
5. Are the students encouraged to use a variety of approaches to solve a problem?
6. Are there assessment exercises that encourage students to estimate their answers and to check their results?
7. Is the science information given in the story problem and elicited in the answer accurate?
8. Is there opportunity for assessing skills (both in the use of science tools and in science thinking) through some exercises that call for hands-on activities?
9. Are there exercises included in the overall assessment strategy that need to be carried out over time?
10. Are there problems with purposely missing or mistaken information that ask students to find errors or critique the way the problem is set up?
11. Are there opportunities for students to make up their own questions/problems or designs?

*These questions are taken from Senta Raizen et al., *Assessment in Elementary Science Education* (Andover, Mass.: The NETWORK, Inc., forthcoming).

solve it. A teacher can observe which students need assistance in this process and what kinds of help are most effective. The levels and types of help needed provide information on each student's competence. Obviously, the teacher needs to construct a series of probes and questions ahead of time that can help students decide how to approach the problem and that will provide the teacher with insights about their learning.

• A teacher can divide students into small groups and assign a science task: classifying a handful of seeds and making observations about the seeds. As the students work, the teacher moves around the room, noting who completes the task easily, who appears to be having difficulty, who has problems using the hand lens and the scales, who contributes interesting ideas, who hangs back, and so on.

Keeping a journal of such observations as the children proceed through several science units in the course of a school year provides rich data on student performance over time. Moreover, the data are in sharp contrast with the "snapshot" provided by a one-time administration of a standardized test. The data collected by the teacher will reflect a student's growth in competencies, rather than a comparison of that youngster's achievement to the achievement of others. In addition, the information comes from several sources that are a better reflection of actual work in science than an artificial multiple-choice test.

To be effective, teachers should observe all children regularly. In that connection, teachers should be aware of the human tendency to note the atypical and neglect the commonplace. Observations of the routine are just as important as observations of the exceptional. These characteristics make the difference between casual observation – which all teachers do all the time – and the systematic observation that provides good data on the performance of all students.

On their face, informal observations seem more valid than more formal, written measures. Unfortunately, on their face they may also appear less reliable. However, reliability comes through replication. Aggregating multiple observations can increase reliability. Validity will be highest when such multiple observations also involve a degree of "convergence," a synthesis of evidence from different contexts, employing different modes of representation. For example, if a child makes drawings illustrating an idea, talks about the idea, and sets up a relevant experiment without having to use any of the probes provided by the teacher, the teacher can assume with confidence that the scientific construct or principle is understood.

There are those who would say that there is little use in attempting to improve the assessment of science until we improve the teaching of science. This view ignores the fact that assessment, through monitoring the outcomes of science instruction, can be a valuable tool to improve science programs. Although interventions are needed in other areas, using assessment can be a powerful strategy for change. By focusing more clearly on the outcomes we wish to achieve, we can effectively change the quality and the quantity of science instruction in elementary schools. Teaching and assessment go hand in hand: change one and the other changes, as well.

Furthermore, by emphasizing the importance of assessment, we stress the importance of science. If the intended learning outcomes – science knowledge, science skills, and the ability to use science in one's life – are important, then the subject itself is a critical component of the curriculum. And assessing the effectiveness of the science curriculum and of science instruction will provide continuing guidance to help us succeed with all our students in all their subjects.

---

1. Rolf Blank and Pamela Espenshade, *State Education Indicators on Science and Mathematics* (Washington, D.C.: Council of Chief State School Officers, 1988).
2. Ina V. S. Mullis and Lynn B. Jenkins, *The Science Report Card: Elements of Risk and Recovery* (Princeton, N.J.: Educational Testing Service, NAEP Report No. 17-S-01, 1988).
3. International Association for the Evaluation of Educational Achievement, *Science Achievement in Seventeen Countries: A Preliminary Report* (Elmsford, N.Y.: Pergamon Press, 1988).
4. Richard J. Murnane and Senta A. Raizen, eds., *Improving Indicators of the Quality of Science and Mathematics Education in Grades K-12* (Washington, D.C.: National Academy Press, 1988), p. 30.
5. Donald W. Dorr-Bremme and Joan L. Herman, *Assessing Student Achievement: A Profile of Classroom Practices* (Los Angeles: UCLA Center for the Study of Evaluation, 1986).
6. Benjamin S. Bloom, "The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring," *Educational Researcher*, vol. 13, 1984, pp. 4-16.
7. Iris S. Weiss, *Report of the 1985-1986 National Survey of Science and Mathematics Education* (Washington, D.C.: National Science Foundation, No. SPE-8317070, 1987), p. 12.
8. Ibid., pp. 64, 67, 91-92.
9. Dorr-Bremme and Herman, op. cit.
10. International Association for the Evaluation of Educational Achievement, op. cit.
11. New York State Education Department, *New York State Elementary Science Program Evaluation Test: Information on Test Administration* (Albany: State University of New York, 1988).
12. *Science for All Americans* (Washington, D.C.: American Association for the Advancement of Science, 1989). K

Reprinted with permission from The Executive Educator, July.
Copyright 1990 the National School Boards Association. All rights reserved.

A changing world requires changes in math instruction

BY GAIL MARSHALL

WHEN IT COMES to modern-day teaching and testing, the United States could take a lesson from ancient China. The ancient Chinese devised tests that could zero in on the finest candidates for civil service jobs. No one expected the finalists simply to know about horsemanship or ceremonial etiquette; they expected candidates to demonstrate proficiency in those skills. There were no multiple-choice tests. Aspiring government clerks had to write, ride, and recite. And they had to do those things well.

As Chinese life changed, civil service tests changed to reflect the new conditions. Riding expertise was no longer required once court clerks lived and worked in government enclaves. And when recording official documents gained major importance, penmanship became the prized performance standard.

The moral of the story: Teaching and testing must stay true to society's needs. It's a critical lesson that the Chinese learned centuries ago, and one the United States must take to heart—especially in mathematics and computer instruction. We can no longer behave as though we are educating future workers for the assembly line.

U.S. life is changing. A technological society is replacing the community of farms and factories. It is becoming more important to adapt to constant change than to perform repetitive actions. In the 21st century, U.S. workers must meet challenges and solve problems.

The National Council of Teachers of Mathematics (NCTM) sees the changes ahead. In its *Curriculum and Evaluation Standards for School Mathematics*, NCTM mandates what and how mathematics should be taught, as well as how performance should be measured to ensure a work force skilled to meet tomorrow's job requirements.

Like changes in Chinese testing, the standards signal a reinterpretation of job performance skills and standards. NCTM recommends a shift decreasing drill and reliance on multiple-choice tests and increasing problem solving and real-world applications.

Getting more from computers

To achieve the new balance, schools must change the way they use computers. The days of passive button pushing to solve multiple-choice problems in drill-and-practice programs are over. NCTM suggests using computers as a key part of the mathematics curriculum, and that calls for a radical reorientation.

Research by Henry Jay Becker at Johns Hopkins University shows that the majority of schools, especially urban schools, use computers for drill. Drill asks for one and only one correct answer. Drill keeps track of correct answers, but it doesn't diagnose the process students use to generate incorrect answers.

NCTM is concerned about the way problems are solved. Counting on fingers to solve a problem isn't as beneficial as learning strategies based on a knowledge of the relation between numbers. Tomorrow's workers, asked to estimate the week's production, shouldn't have to say, "Wait a minute. On Monday we made 251 widgets and on Tuesday we made 252, so 251 plus 2 is 253 and 253 plus 50 is 303 and 303 plus 200 is 503." They need estimation strategies that are less laborious and more efficient. They must develop accurate methods based on a richer understanding of mathematics than our current drill yields.

NCTM stresses real-world activities. Simulation software is an excellent way to provide mathematical challenges. Simulations are "micro worlds" that model complex real-world events on a computer screen. Simulation software lets students run a business, calculate profits and losses, estimate inventories, and hypothesize reasons for success or failure. It lets them simulate a space trip, making complex decisions, such as how to coordinate food, fuel, and equipment while remaining mindful of costs. Whenever students work with simulation software, they're constructing mathematical models to explain events.

Problem solving, another NCTM recommendation, becomes more dynamic in the computer classroom. Many teachers enjoy teaching problem solving. But problem-solving activities in textbooks aren't mathematically rich, and teachers often lack time to devise their own. Other teachers emphasize drill because they have limited mathematical backgrounds and lack confidence to generate problems on their own.

Computer-based problem solving provides novel problems, many of them in real-world guise. The computer's ability to create "new" problems each time makes it an ideal tool for developing students' understanding of difficult concepts. Each time students use the computer, it generates a new set of conditions, and they must develop strategies to work with the new conditions. Because the computer has been programmed for the best strategies and answers, students get immediate feedback.

The computer won't automatically display the correct answer (or it wouldn't be problem solving). Instead, the computer provides suggestions. The open-endedness of the task challenges students to search for alternatives, combine several strategies, and collaborate. Those are the skills transforming the workplace, and those are the skills that belong in class.

The visual cues, animation, and suggestions built into many programs make the computer an ideal partner in problem solving. Many problems are ambiguous.

Gail Marshall is an education consultant, software developer, and researcher in St. Louis.

BEST COPY AVAILABLE

just like real life. Many problems can be solved more than one way. That stimulates students to look closely at a problem and evaluate it, just as they must do in everyday situations.

Other necessary changes

But using simulation or problem-solving software isn't going to have the impact NCTM expects if students work alone, as they do in many classrooms. The culture of the American workplace is changing. Americans must realize that the assembly-line model of education, with each student working alone toward mastery of an isolated task, doesn't fit industry's needs for the 21st century. Collaboration is the key, both on the job and in the classroom. Problem solving thrives in cooperative settings.

Schools must rearrange classrooms to accommodate problem solving and simulations. If there's only one computer, it must be moved front and center. Teachers need a large-screen television monitor so they can coach a group of students or the entire class. When the computer isn't hooked up to a television monitor, it should be located where five, six, or eight students can work together.

The computer lab, with its one-computer-to-one-student setup, is too restrictive for simulations and problem solving. Laboratory computers should be moved into the hall or library so students can work collaboratively. If the computers remain in the lab, teachers should offer students additional tools, such as almanacs and atlases.

Also, schools must devote a day, a week, or a month to simulation and problem solving. A good simulation calls for collecting and analyzing computer data. That takes time. Teachers can't (and shouldn't) turn off the computer just because the schedule calls for social studies to follow math. If chosen carefully, simulation and problem-solving software will also apply to language arts, science, and social studies.

But responding to the NCTM standards means more than rewriting a school system's objectives, changing textbooks, moving computers, or rearranging furniture. If computers are going to be used as NCTM envisions, schools must support training in different teaching and learning strategies. Alternative assessments, based on observing students as they work with novel problems, must replace multiple-choice tests. And administrators must allow time for teachers to try and modify new instructional strategies.

Many teachers who rely on drill programs will find it difficult to switch gears. But by allowing drill to dominate instruction, they've stopped worrying about analytic thinking, evaluation of options, or the synthesis of different ideas. The computer has become the curriculum instead of being a curricular tool.

Achievement tests, models for content, and the format of computerized drills must also change. Critics of achievement tests—such as NCTM, state departments of education, elected officials, and teacher organizations—have called for alternative assessments. Just as the ancient Chinese watched a horseman to see how well he could ride, math teachers must watch a student solve probability problems. Just as civil servants in ancient China had to have graceful handwriting, students should have the ability to develop a hypothesis encompassing all the data.

Ancient China was able to make testing changes because its people agreed on cultural values. Unanimity about American education goals doesn't exist. Drill still exerts a powerful hold. But going backward to the old basics of counting and computation is no longer an option. We need to go forward if American workers are going to compete successfully in the technological and political revolutions occurring worldwide. ■

Reprinted with permission from English Journal, February 1992.

A Large-Scale Assessment to Support the Process Paradigm

Charlotte W. O'Brien

Most writing teachers accept the principle of accountability. They understand the need to show what progress students are making in learning to write. However, they have learned that large-scale assessments often violate the process-centered approaches supported by research, their own classroom experiences, and workshops such as those affiliated with the National Writing Project.

Because of the instructional implications, most writing teachers agree that students' writing ability should be assessed through actual writing samples (direct measures) and not by multiple-choice questions about someone else's writing (indirect measures). Even so, assessment is not without its problems. The major difficulty is that a single writing sample produced in a limited time period is an unfair measure of writing ability and yields little or no indication of students' understanding of the processes of writing. This kind of assessment, as Willa Wolcott says, works at "cross-purposes with current writing-instruction theories, severely curtailing—especially for weaker writers—any potential applications of the process paradigm" (1987, 40).

For too long, direct writing assessments have been based on writing samples collected from students who are given a writing prompt and expected to write a proficient essay under time constraints ranging from fifteen to fifty minutes. Indeed, assessing student writing through an impromptu essay sends the message to developing writers that they do not need time to think and work at their writing and that revision is not important. Furthermore, students usually have no idea what the criteria for grading are, and, in many cases, neither do their teachers. In fact, many teachers do not feel compelled to prepare their students in the writing-process approach because they know students will not be assessed that way.

As researchers continue to observe, categorize, and report what writers actually do when they write and as the writing projects nationwide make teachers more aware of themselves as writers, many educators have begun to feel uncomfortable with writing assessments that contradict the way writing is being taught in the classroom. This is not to say that writers never need to write impromptu memos, letters, and answers to examination questions, but large-scale writing assessments, because of the impact they can have on instruction, should reflect current composition pedagogy.

Posing the Question

Up until last year writing had been assessed in Missouri through a random sampling of students in grades six, eight, and ten. These students all wrote to the same prompt for one class period. During the summer, teachers trained as raters scored the papers on a six-point holistic scale. Predictably, this type of assessment had little or no effect on the teaching and learning of writing across the state, except perhaps for the thirty teachers involved in the scoring session. After much discussion about how the state assessment might have more of an impact on instruction, an advisory committee of writing teachers, directors of state writing projects, and curriculum consult-

ants, with the support of the commissioner of education, ventured to ask this question: Could the statewide assessment be designed to satisfy assessment requirements and at the same time to support process-writing instruction?

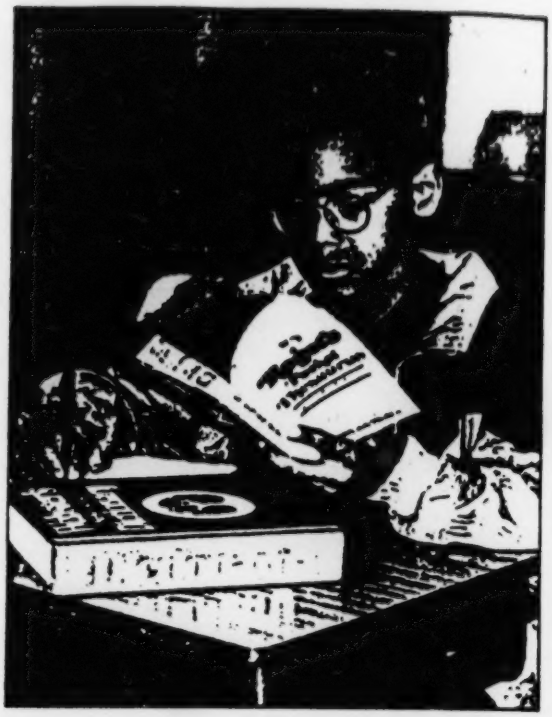
At least one research study showed that it could be done. On a large-scale direct writing assessment of thirty-six classes of fourth graders randomly assigned to traditional or process groups, Mary Ellen Farmer found that the quality of writing was higher for the process group and that use of a process approach did not affect test validity any more than the traditional approach did (1989). These findings were corroborated by a pilot study conducted in conjunction with the field testing of a prompt for our state assessment. Eighth graders from rural, suburban, and urban schools were assigned to process or traditional (impromptu) groups and wrote for either three 45-minute periods using a writing workshop format or for one 45-minute period culminating in a "product" (O'Brien 1991). Both groups responded to the same prompt. The scores of students completing the process task were significantly higher than those of students completing the impromptu task. These studies convinced us that a process task was a more equitable method of assessing writing at the eighth-grade level. To assess 54,000 eighth graders in Missouri, we would use a process format.

Designing a Process Writing Assessment

Test developers are very suspicious of any type of assessment that is not standardized. Traditionalists view writing as a solitary task and thus easily tested; however, advocates of the process approach see writing as a collaborative journey, in which the author is ultimately responsible for making the decisions that shape the piece of writing. Our challenge was to develop a process-writing assessment that would be standardized enough to fulfill testing requirements yet allow the flexibility needed for a writing-workshop format.

We wanted the assessment to serve as a model for incorporating process writing into the classroom.

From the field test, we had a good idea of which



areas needed to be refined and which seemed to work well. Most important, we wanted the assessment to serve as a model for incorporating process writing into the classroom. We wanted teachers to view it not as an intrusion on their class time but as a teaching unit similar to what they would be doing anyway. Because our primary objective was to encourage and support process-writing instruction rather than to test skills on various types of writing, we decided to use a prompt that would elicit a personal narrative response.

Agreeing with Wolcott that there is a need for more interplay between process and product in assessments, we recognized with Marie J. Lederman the importance of product.

> The best teachers do help students learn something about their own writing processes. . . . But in the end, it is a lie to tell students that "product" does not matter. . . . In the real world, product is all we can share with each other. (1986, 41)

In a process-oriented assessment students are invited to use process strategies and receive feedback from their peers while they are writing, but they also know that their written product will be rated by teachers who do not know them.

Developing a Process-Writing Task

All public-school students in Missouri taking eighth-grade English participated in the assessment, which encompassed three 45-minute blocks of time on consecutive days. Teachers were to serve as facilitators rather than as instructors. They read aloud brief scripted instructions which advised students as to the flexibility of the time frame and suggested workshop activities. Students responded to the following prompt:

> Most of us have had to attend events that we didn't want to attend. Even though we tried all kinds of excuses, we still had to go. Write a paper describing an experience in which you had to attend an event that you didn't want to attend. Include your thoughts and feelings about the experience. (Missouri Writing Assessment 1990, 1)

For ten to twenty minutes on the first day, students had time for a variety of prewriting activities (for example, drawing, listing, clustering, mapping, free-writing) which could also include collaborative brainstorming and sharing in large or small groups. When they were ready, they began drafting their compositions. On the second day, they continued drafting and revising their papers and engaged in a variety of collaborative revision activities, including peer workshops. The following questions guided them in reviewing their own and their classmates' drafts:

1. Does my paper stay focused on the controlling ideas?
2. Does the order of my paper make sense?
3. Have I started my paper in an interesting way?
4. Does my paper have a satisfying ending?
5. Have I used description/details/examples?
6. Have I carefully edited my paper for spelling, punctuation, and grammar errors? (4)

If the answer to any of these questions was "No," teachers reminded students to revise their papers so the answer would be "Yes." Students were free to use any references or usage handbooks found in the classroom.

Each student received two 4-paged booklets. For two class periods students wrote in Booklet 1, which contained the prompt, space for prewriting, and three pages of lined paper with plenty of margin space for making additions and revisions. The questions to guide peer discussions were printed at the bottom of the last page. On the final day, in addition to Booklet 1, students received Booklet 2, which contained a page for demographic information and space for writing the final draft in ink.

The format worked well. Even teachers who said they did not normally teach writing using a process approach said they had no problem following the instructions, and many were surprised that their students used all of the time allotted to write their papers. We encouraged teachers to respond to the prompt along with their students, but few did. Several said that, because neither they nor their students were used to this type of assessment, they felt they had to spend the time monitoring their students' progress rather than writing themselves.

Developing the Scoring Criteria and Selecting Range-Finders

To fine-tune the scoring criteria (rubric) we studied the papers written in response to the field-tested version of the prompt. For the actual assessment we wanted to be sure that the rubric really reflected the characteristics of these papers. One important change we made was to write the descriptions of the score points in paragraph form. We had found during the field test that listing the characteristics separately made raters focus on specific features rather than on the papers as a whole. Eighth-grade teachers received a copy of the revised rubric before they gave the writing assessment and were encouraged to share it with their students.

A committee of twenty teachers and consultants conducted what we called the "scoring Olympics" to select range-finders for each score point from the papers written during the statewide assessment. We were trained in holistic scoring procedures using the rubric and range-finders selected from the field-tested papers. Then each of us, working individually, read randomly selected papers from different geographical areas in the state, choosing ten to fifteen which we felt might be good anchors. Later that afternoon we met together in a large circle. We read the paper under consideration simultaneously and held up a card to indicate our rating. A paper had to receive at least 70% agreement from the committee to be considered for use in the training packets. The range-finders (anchors) had to have 90-100% agreement.

From the eighty to ninety papers selected that day, we developed training, practice, and certification packets to use in training Missouri teachers to

score the assessment. The training packets contained range-finders for each score point. Practice and certification packets each contained ten papers representing any number of score points. Some of these papers fit the descriptions on the rubric perfectly; others were problem papers containing characteristics of several score points and were included for discussion. Many of these papers fell in the mid-range—the area raters have the most difficulty scoring.

Standardizing Training and Scoring Procedures

The eighth graders had done their part. Papers from the Ozarks to the Boot Heel, from St. Joseph to St. Louis, from every urban, suburban, and rural community across the state were stacked in boxes in the warehouse. Suddenly 54,000 became a real number to me, not just a number on a page. I woke up many a night to find myself multiplying and dividing, trying to figure how long it would take how many raters to score how many papers.

At times, however, it seemed that the logistics of scoring the papers were the least of my worries. We had limited funding for the project. There was enough to pay for substitute teachers for only one day. That meant we had to ask 450 teachers to give up a Saturday to help us score the papers. Amazingly enough, they were willing to do it, and some of them did it several times. I was able to recruit twelve site coordinators from our writing-advisory committee. (We later had to add another two sites because of cancellations and bad weather.) Coordinators chose sites in the areas of the state where they lived, made arrangements for the scoring, and recruited thirty to thirty-five teachers each to serve as raters.

The problem was not how to how to get the papers read but how to train raters to be consistent at fourteen widely separated sites.

This scoring took place on consecutive weekends in January and February lasting Thursday evening through Saturday afternoon. Each site scored approximately four-thousand papers from another geographic area of the state. A different

group of three sites scored simultaneously each week. We knew we would be up against unpredictable Missouri winter, but sometimes the bad weather worked in our favor. For example, in Neosho, one of the first sites, the schools had been closed for two weeks with 6-8 inches of ice covering the ground on most of the rural school-bus routes. Those teachers were looking for any excuses to get out of the house, and we were happy to entertain them. A few weeks later in St. Joseph it snowed ten inches the first day of the scoring. We wondered if anyone would come back the second day. They not only came roaring through the snow in their four-wheel drive vehicles, but even had time to design sweat shirts, commemorating the event with the slogan "Raters of the Lost Art."

Basically, the problem was not how to get the papers read but how to train raters to be consistent at fourteen widely separated scoring sites in the state. We wanted a supportive and nonthreatening atmosphere where professionals would feel free to discuss why certain papers did or did not reflect the score points as stated in the rubric, yet be able to reach the consensus necessary for scoring.

Training was thorough and practice was plen-

tiful before the scoring actually began. Because many of our raters had never been exposed to holistic scoring, we began on Thursday evening with a discussion of the philosophy, procedures, and uses. The trainer used the range-finders for discussion and training. Then on Friday morning after a review of the rubric and range-finders, raters scored two packets of ten papers each, compared their scores with those of the committee, and discussed any discrepancies. The training took four to five hours depending on the experience of the raters. Before beginning to score the actual papers from the assessment, raters had to qualify by scoring a certification packet. They were required to score eighty percent of the papers within one score point of the previously agreed upon score, with fifty percent right on target. Many were nervous about doing this, but all felt that it was necessary to maintain scoring consistency. To their surprise, the majority found themselves scoring one-hundred percent in agreement. Those who had trouble meeting the standard reviewed with the trainer the papers they had scored discrepantly. Then they scored a different certification packet before beginning to rate the assessment papers. Usually the trainer found that the rater was scoring consistently higher or consistently lower than the committee. Once that was pointed out, the rater usually had no trouble scoring on target.

Reporting the Results of the Assessment

Each building and district received a group summary report which presented the results of the assessment, indicating the number and percentage at each score point for their students as well as for all students in the state. A copy of the rubric was attached to the report. Such data served as a point for districts to begin assessing their current writing programs.

Teachers who had served as raters for earlier state samples which had been written under the traditional one-class-period format indicated that they felt the quality of the papers in this assessment was much improved. Rather than a bell-shaped curve, the statewide curve was negatively skewed with approximately sixty-four percent of the students scoring in the upper half of the scale.

Building a Statewide Network of Writing Teachers

As writing teachers, how could we rate student writing and not write something ourselves? On the last day of the scoring session participants responded with a free-writing. Many agreed with a first-year teacher who commented that taking part in the assessment forced her to clarify her own views on "why it is that we teach (and require) children to write essays in the first place." Others wrote of the professionalism they had observed among their colleagues and of their support for a statewide writing assessment which actually reflected the way writing was being taught in their classrooms. In a lighter vein, this exhausted rater described the experience tongue-in-cheek.

We had also built, through the common experience of scoring the papers, a network of writing teachers.

> I pause to reflect upon the tattered remnants of my mental awareness—numbed by sense overload, short-circuited by variant currents of thought. For hours, an endless drone of haphazardly written words has bombarded my mind. I am left a saturated

> sponge—dripping with both the pure and the polluted, the fresh and the tainted liquid of human thought.

Of course, some of the raters did not know exactly what they were getting into, but most of them, when asked if they would participate again, enthusiastically replied, "Yes!"

Reflecting that the scoring experience was the reminder she needed, a St. Louis teacher wrote,

> I teach language because I admire and respect the others who do so, and I teach early adolescents because I take so much delight in the things they say, do, and write.

Many of us enjoyed the opportunity to see into the eighth-grade psyche. We concurred with the junior-high teacher from Cape Girardeau who wrote:

> For two days I have been transported back to glimpse family reunions, weddings, summer camps, piano recitals, junior high dances, and visits to Grandmother's house from their young perspectives. The papers all give evidence that our students are alive and well, coming of age, bearing those same initiation rites other generations before them have borne. Some papers have sparkled with creativity from beginning to end. Others have been more ordinary—with the exception of a wonderful line or phrase, so honestly turned, so uniquely fresh, so unreservedly offered. . . .
>
> The reading has also been bittersweet. For there have been those papers about Dad in prison, Mom wanting to be free, lost friends, the traumas of forced moves, dear relatives' funerals, and "alert" papers on experiences which changed lives forever. It has never been easy growing up!

But most of all, teachers commented on the camaraderie and strength of purpose which developed as we tackled the seemingly insurmountable task. No one expressed it more vividly than this teacher:

> Our table bonded as the piles of papers dwindled, and by lunch on Saturday, we were thinking of ourselves as the Bad-Ass Holistic Scorers from Hell!

We had not only made a good start in (1) supporting the teaching and learning of process writing across the state, (2) letting students know that writing is an academic skill for which they will be held accountable, and (3) providing educators with an overview of eighth-grade students' writing ability. We had also built, through the common experience of scoring the papers, a network of writing teachers far beyond school district boundaries. Most important of all, we had made connections—with ourselves as writing teachers, with our students as writers, and with our methods of teaching and assessing writing.

Missouri Department of Elementary and Secondary Education
Jefferson City, Missouri 65102

Works Cited

Farmer, Mary Ellen. 1988. "A Comparison of Process and Traditional Approaches to Large-Scale Writing Assessment: Investigating Score Magnitude, Interrater Reliability and Concurrent Validity." Diss. Catholic U.

Lederman, Marie J. 1986. "Why Test?" *Writing Assessment: Issues and Strategies*. Ed. Karen L. Greenberg, Harvey S. Weiner, and Richard A. Donovan. New York: Longman. 35–43.

Missouri Writing Assessment. 1990. *Eighth Grade Test Booklet 1*. Jefferson City: Missouri Department of Education.

O'Brien, Charlotte. 1991. "A Comparison of an Impromptu and a Process Task on a Large-Scale Direct Writing Assessment of Eighth Graders." Diss. U of Missouri.

Wolcott, Willa. 1987. "Writing Instruction and Assessment: The Need for Interplay Between Process and Product." *College Composition and Communication* 38 (Feb.): 40–46.

SECTION 3

A LOOK AT PROGRAMS

Using Writing Portfolios:

Reprinted with permission of ECS Learning Systems, Inc., PO Box 791437, San Antonio, TX 78279, 512-438-4262. This interview was conducted by Lori Mammen.

An interview with John Collins

Editor's Note: Interview With An Authority is a regular feature of WRITING TEACHER. In this issue we focus on our recent conversation with Dr. John Collins of The NETWORK, Inc., Andover, Massachusetts. Dr. Collins spends his professional life designing programs and conducting in-service training in writing for language arts and content area teachers. His major focus is the design of programs that convert research on writing into practical and time-saving teaching techniques. In this interview Dr. Collins discusses the use of writing portfolios to improve the quality of student writing.

WT: What role does evaluation play in the development of effecve writing skills? Do some evaluation methods work better than others?

COLLINS: Let me set the context for how I would like to answer these questions. I've been doing a great deal of work lately in writing across the curriculum and working with teachers who are not primarily English teachers. What I've been trying to do with these teachers is talk about different types of writing.

I talk about five types of writing and this gives us a way to talk about evaluation. For Type One writing, the student writes something and the writing gets a check. If the student doesn't write the requisite amount, the writing gets a minus. It is quite simple; the evaluation is only to get the kid's head off the desk. We talk about using Type One writing in place of a class discussion, which highly verbal kids often dominate. Type One writing is one way that we can get everyone thinking about the topics we are teaching. What do the students know about the pyramids? What do the students remember about the periodic table? Type One writing probably should not be evaluated because the point is simply to try and "check" inside someone's head, or to let students go through their own mental files for information.

For Type Two writing, the teacher only worries about correct content. This type of writing is an attempt to offer a writing alternative to non-writing teachers in lieu of multiple-choice tests and worksheets so that the students have to construct some information. If teachers label writing as Type Two, this gives them permission not to correct the mechanics, the punctuation, the spelling.

Type Three writing has focus correction areas (FCAs). This is when teachers begin to become serious about evaluation because they are looking at writing skills as well as the ideas or the content. Focus correction areas can include things that relate to content (ten facts, a thesis statement) to style (sentence variety, organizational structure), to mechanics and conventions (end punctuation, spelling), or to any combination of the three. The evaluation for Type Three writing is related directly to the FCAs assigned to a given piece of writing.

With Type Three writing, students begin to really learn how to write. As an editor, if you sit down with a manuscript that I wrote and you make all the corrections and changes, then I am probably not going to learn how to write from the experience because I will never write this piece again. On the other hand, if you say, "John, I looked at this piece of writing that you did, and you seem to have three major problems. Your paragraphs are too short, your sentences are too long, and I think you need more details to support your general statements." Then we would talk about each of these focus correction areas, and I could learn a great deal from the experience that could be applied to my future work.

WT: Which teachers should be using Type Three writing? Is it mainly for the English/language arts teachers?

COLLINS: When I work with teachers, I try to convince everyone of the benefits of Type Three writing. This week I was working with a group of middle school teachers who work with students in clus-

BEST COPY AVAILABLE 57

ters. One particular cluster has developed a marvelous use of the focus correction areas. The English and reading teachers looked at their students' writing during September and diagnosed the weaknesses in the writing. Then they reported to the cluster and identified the weak areas in the students' writing: insufficient support for general statements and the need for more complex and varied sentence structure. They identified these areas as their FCAs. The other teachers (science, math, etc.) were not asked to teach these skills. However, when the English and reading teachers finished teaching these skills, the other content area teachers were asked to require Type Three writing that had these FCAs. I have worked with a number of school systems that have used this method, and the other content area teachers find it very palatable, especially middle school teachers who theoretically are interested in the "whole kid" in terms of basic skills. Teachers become very excited about the approach, but they are particularly excited that they do not have to become writing teachers.

"The major weakness in student writing in the United States today is the students' inability to improve a first draft."

WT: Which is the greatest fear of the other content teachers, isn't it?

COLLINS: We probably don't even need to discuss that point. They are worried about their competency and their own writing skills. But imagine you are a science teacher and I say to you, "Lori, would you give writing assignments that require students to include supporting details for general statements and sentence variety?" Imagine, too, that I have explained to you the definition of those FCAs and how I had taught it. I would only ask you to grade the writing for those FCAs and your own focus criteria such as the correctness of the information included in the writing. You could even allocate points for each of those FCAs so students receive specific feedback.

A student might receive thirty points for supporting general statements with specific details, fifty points for the correctness of the information, and twenty points for sentence variety.

WT: Does this grade become a science grade, a composition grade, or both?

COLLINS: It becomes a science grade. We have found that it is too complex to handle the grading in any other way. Also, teachers were invading each other's territory.

WT: Of course, by counting it as the science grade, students also see the connection between what they are doing in English class and science class.

COLLINS: Yes, and we joke all the time about how the kids hate it because all the teachers have gotten their "wagons in a circle."

WT: Is there any other important element of Type Three writing?

COLLINS: After a student has finished a Type Three writing, the teacher allows five minutes for the students to read their pieces aloud simultaneously. In other words, everyone is reading, but no individual can hear what another has written because they are all reading in a one-foot voice. Oral reading is a critical part of Type Three writing.

WT: Why do the students need to hear their own voices reading their writing?

COLLINS: This, to me, is a critical element of my program. People remember focus, correcting, and I am introduced with, "Here is John Collins, who is going to do a workshop on focus correcting. I know you've all been waiting to hear about that." Then my life passes in front of me! Yesterday, I encountered a principal who, after an in-service session, wrote an article for the local newspaper about my program. Throughout the entire article he talked about focus correcting. After I read it, I said to him, "Where's oral reading?" He said, "Oh, my! I forgot about that!" That is my worst fear.

All the research I read says that most kids can write first drafts. They are not non-fluent. This is one of the many accomplishments of the writing process movement over the last twenty years. Kids may be more willing to write than they used to be. I see this when I do model lessons in classrooms. I have a very easy time getting kids to pick up a pencil or pen. The major weakness in student writing in the United States today is the students' inability to improve a first draft. The teachers I talk to all agree that this is a major problem; I have never yet had a teacher disagree with that.

The single most important thing for a student to improve a first draft is to hear it. They must literally listen to it. This requires skill and repeated practice. If a teacher has students do this only three or four times per school year, they will not do it the following year for their next teacher. The teacher must institutionalize the process in the classroom, encouraging the students to follow it in a consistent way. And for a school building to make a real, long-term difference in the way students write, then teachers should follow the same process. By the third year, most students will do it almost as an involuntary muscle reflex because they wouldn't think of

submitting a paper that they had not read aloud. They have seen how many mistakes they make.

WT: Then this oral reading leads to some other type of writing?

COLLINS: Type Four and Type Five. Type Four is the basis of all my work. It expands on Type Three by adding peer editing, but it still asks the teacher to limit negative feedback to the FCAs. Type Five writing is publishable writing. It is the type of writing you want in *Writing Teacher*; it is what I want in my books. When I work with a group of teachers, my belief is that the teachers sitting in front of me want to give writing assignments, and they want them to be Type Five – publishable writing, "perfect" writing. This writing can stand alone without any explanation, qualification, or apology.

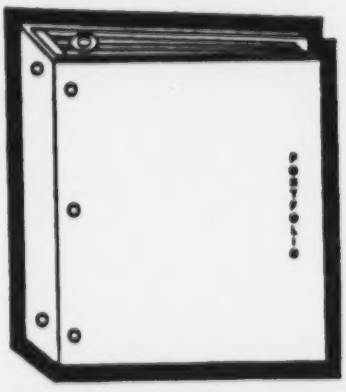
This is a long prelude to the question about evaluation, but I think the evaluation question has a great deal to do with the type of writing a teacher is asking a student to do. If a teacher has very good students, they can do excellent work. The teacher should demand that they produce excellence – more Type Five work. But it is not the most efficient way to teach writing. Type Three and Type Four are the writing assignments that are the best way to teach writing skills.

WT: I think I hear you saying that in some classrooms only one or two types of writing occur. So certain kinds of evaluation never happen, or the wrong kind of evaluation occurs on some pieces of writing.

COLLINS: Exactly. I have some rules for assigning Type Five writing. If teachers can answer "yes" to two of four questions, then they should assign Type Five writing. First, is the writing going to be read to an audience outside the classroom? If it is, then it is a Type Five. Second, are the students sufficiently motivated to do Type Five writing? Third, are they sufficiently skilled to do Type Five writing? Non-writers can suddenly hit the eighth grade, and hear the eighth grade teacher say, "You're going to do a research paper, and it is going to be perfect." This is too big a leap for the students so teachers receive plagiarized papers. The kids have no other choice. Fourth and finally, are there time and conditions created to make it possible to do Type Five writing? Is there time for oral reading, peer and teacher feedback and evaluation, and multiple drafts? If you have the time in your curriculum, then you can make it happen. If a teacher is just going to give an assignment and then walk away, it is not fair to the kids. This is assigning writing, not teaching writing. Unfortunately, this happens to students all the time.

WT: Isn't this how many of us experienced writing in school?

COLLINS: Yes, this is how I experienced writing when I was in school. All my teachers did was "assign" writing. If we didn't know how to write, we didn't even know what we didn't know. In a situation like this, a student doesn't even know if there are patterns of errors in the writing. For example, you could write awkward sentences for the rest of your life. They will never sound as awkward to you because you wrote them. This is not because you are stupid, or because you cannot write. If you tend to have a lot of awkward sentences in everything you write, it would be helpful for you to know why. Maybe it's because you're trying to get too many ideas in one sentence, or you never use dashes that would help break the thoughts apart. Wouldn't it be nice if you could walk away with some learning from this? Maybe you could learn how to shorten sentences so you wouldn't get into these long, involved ideas that you or your reader can't handle.



WT: How do writing portfolios fit into this idea of teaching writing versus assigning writing?

COLLINS: I want to emphasize the notion of letting students know what the criteria for judging a paper are going to be. First, they must know what type of writing they are doing. I stress with teachers that if they limit their feedback to a few criteria, then they can evaluate those papers more clearly, and they can be better writing coaches for their students. A good athletic coach does not tell the athlete everything that was done wrong on a play, but tries to be specific about the feedback.

WT: How do teachers go about choosing the criteria they will use for evaluation of student papers?

COLLINS: The choice of criteria is the most difficult part of the program, and you've asked the hardest question you could ask. In the Cumulative Writing Folder™ program, I have a page of focus correction areas for different domains of writing. Writing is divided into four domains: imaginative-narrative, practical-informative, sensory-descriptive, analytical-expository. Below each category of writing are listed some FCAs that might work for each category. You probably are an extraordinarily competent journalistic writer. You probably are very concise and very good at organization because you know you have to present information in a systematic way. But, if I asked you to write a description of your backyard in the spring, you might be at a loss. You don't do sensory-descriptive writing. On the other hand, your husband, who is a poet, might be terrific at describing the backyard, but be less effective writing up an interview with John Collins. I believe these specific skills are acquired. We can learn, almost unconsciously, the forms, formats, and strategies to get certain writing jobs done. I am sure that your tenth interview for *Writing Teacher* was much easier to do than your first one.

This is the point – we ask the teachers which of the four domains of writing they want to emphasize during a given school year. A teacher cannot teach a kid how to do all four kinds of writing

BEST COPY AVAILABLE

in one year. That is the biggest mistake many teachers make because no one can teach writing in one year. Within the chosen domain of writing, there are certain primary traits, or FCAs, that are critical for success in that type of writing. For example, practical-informative writing has to be well-organized so the reader does not become lost. Practical-informative writing probably has short, clear sentences and paragraphs. So, the teacher must make two choices. What domain of writing will be emphasized during the school year? What weaknesses do my students have in this kind of writing?

"A good athletic coach does not tell the athlete everything that was done wrong on a play, but tries to be specific about the feedback."

WT: For a program like this to have long-term success, you are looking at cooperation across the grade levels.

COLLINS: During the first year, we tell teachers to find their own way. What inevitably happens is that, when the program catches on and becomes as powerful as we know it can be, then the teachers will begin cooperating across the grade levels. They realize they can send students from their school knowing how to write, but they also know they won't have to do it all by themselves. This is really exciting!

WT: How do you respond to the critics who believe that teachers should be teaching everything, not focusing on a few FCAs? There must be a great deal of PR involved in "selling" this program to parents, for example.

COLLINS: I created this approach as a classroom teacher based on trial and error, my own research, and stealing from the best teachers I knew. When I began doing consulting, I was never going to publish this program because I believed no one could do it except for me. I believed that parents would "eat teachers alive" if they didn't correct every error on every line of a student's paper. I have sold a million and a half Cumulative Writing Folders™. We have a toll-free number for questions and problems, and I have met many of the teachers who are using the program. The one statement I never hear is "Parents are angry at us." It is amazing to me.

You don't have to do a lot of PR. First, I don't think many parents know what is going on. The students can explain the process to the parents if the FCAs are listed in the upper, left-hand corner of the paper. For example, a sixth-grade student may come home with a paper that has obvious flaws in it, but that also has a high grade. Parents might ask, "How did you get such a high grade on this paper? It has mistakes." The student might answer, "Well, it was a paper where I had to identify a type of rock and explain why this type of rock was the type of rock it was. I had to give five facts about the rock, and I had five facts. I had to have sentence variety. Look, I had several compound and complex sentences. I also was supposed to have sensory adjectives, and I had them. So, I got an A."

WT: When the parent hears the student being that articulate about writing, there is little room for criticism, is there?

COLLINS: Right, the parents don't mind. This also lets me bring up another beautiful part of this program. It's my favorite part! Now, I have just given the student a very good grade on that rock paper, but it does have problems. Let's say it is a page and a half long, but it's only one paragraph. The student places the paper in the Cumulative Writing Folder™, and every paper that is put into the folder is numbered. Later when I want to teach a new skill, such as adjectives for sensory descriptive writing, paragraphing, or transitions, I have the task papers for the students. For example, maybe I am doing a rather "traditional" lesson on effective transitions, and the kids are kind of "traditionally" bored. I say, "Everyone, please get your Cumulative Writing Folder™. I would like you to take out composition four, the one about the rock. Could you see if you can find any transitions in your composition?" I think this certainly beats a worksheet! This leads me back to your question on portfolios. The writing folder is not just for storage; it can be used to help students apply new skills to past papers.

WT: Students also learn that even though they may have earned an A on the original assignment...

COLLINS: ... they can still do a lot of things to improve it. I also sell it to teachers in another way. How long did it take me to grade the original paper? No time! When I was still teaching and didn't think my students were "with me," I only had to look in the direction of the file cabinet, and hands would go up. They knew that the situation could turn into a test. "It seems everyone knows what transitions are. Since you don't have any questions, that's terrific! Bill, would you pass out the writing folders, please? Everyone take out paper four and circle any transitional words that you have on that paper. I know you know what transitions are. If you miss a transition on your paper, you get a B; if you miss more than two transitions, you get an F. This counts as a quiz grade. Begin." Suddenly, I had kids waking up all over the room. Of course, I also got kids who said, "I used transitions" as they looked at their papers.

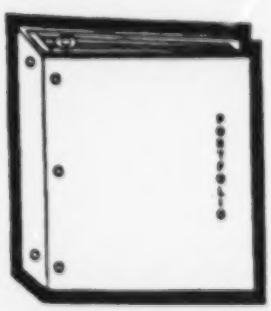
WT: So you validated them.

COLLINS: Exactly – they were validated. Or I got kids who said, "I wrote a page and a half, and there are no transition words except *and*." Then I could ask, "Is that effective writing? Could you make your writing sound more mature? Would it help if you did use *finally, first, second, consequently*, or some other transition word?" I think one of our biggest mistakes in teaching writing is that kids never see a transfer between the workbooks and their own writing. If I haven't corrected all the errors on students' papers, then those papers can become worksheets in the future.

WT: What other benefits are there to using writing folders with students?

COLLINS: My writing portfolio program is different than many other portfolio programs because it is so very organized. One of my concerns about portfolio evaluation is that there are teachers who have students place everything in their portfolios. So that some portfolios become something like "me books;" but I wonder when these kids are taught how to write.

My portfolios contain Type Four writing. Everything that goes into the portfolios has been revised and edited by another student, and it has been critiqued or evaluated by the teacher. Only then does it go into the Cumulative Writing Folder™. Then it becomes a potential worksheet for later. When I am looking at a student's portfolio, I am not reading notes or boyfriend letters. This is more serious writing so I can evaluate the weaknesses of the writer much differently. One of the things that concerns me is that the people who are evaluating portfolios know what is in the portfolios and why.



WT: Each paper in the writing folder lists the FCAs that were evaluated so the teacher can go back and recall what was emphasized on each assignment.

COLLINS: Yes, they can go back and evaluate. Teachers can say, "This student has covered twenty skills this year, but she doesn't seem to know how to do them." On the other hand, another student may have covered only six skills, but may show real growth in all six. This is very exciting to me.

WT: So, the greatest benefit seems to be that this portfolio becomes a real tool for teaching.

COLLINS: Yes. We try to get an entire building or school system to teach the mechanics of the portfolio during one school year. Then in the future teachers do not have to worry that the students don't know what is happening with the portfolios. For example, my students would say to me, "Is this going into the folder?" If I answered, "Yes," then they knew that another student was going to read it. They knew not to "pour their hearts out" about private ideas unless they wanted another student to see it. They also knew that if it was going in the writing folder, someone might come into the room and look in the folder.

WT: Many teachers will want to know what happens to the Type One and Type Two writing that does not go into the writing folder.

COLLINS: You can do whatever you want with it. Much of it will get thrown away. Some teachers like to keep it. Many Type One and Type Two writings are the notes or prewriting experiences that evolve into Type Three and Type Four writings. This prewriting may be attached to the Type Three and Type Four writings, but my personal preference is to throw the drafts away. I found that keeping everything became messy and unwieldy.

WT: So the students come to realize that the portfolio contains only certain types of writing.

COLLINS: Yes, and it gives a way to talk about writing. A teacher or student can say, "What type of writing is this?" I make students label each paper. If it is a Type One writing assignment, then the student knows that it does not have to be perfect. Type One is primarily a fluency activity. It's very much like a journal or a log although I try not to use those terms anymore because we have had so many problems with journals.

WT: How do teachers get started if they are interested in beginning a portfolio program in the classroom?

COLLINS: First, teachers must ask themselves what domain of writing they want to emphasize during a school year – practical, narrative, descriptive, or analytical. That doesn't mean their students will only do that domain of writing, but it will be the kind of writing that will be emphasized. Secondly, the teachers must determine which skills their students will need to do that kind of writing. Then the teachers create assignments – and I am a great believer in giving assignments to students – that would help produce the kind of writing that will be emphasized. As the grade levels go up, I do believe that students should be writing more and more about what they are reading and learning in their classes. If the teacher has been doing a lot of work on figures of speech because it is part of his/her reading or literature curriculum and is emphasizing sensory-descriptive writing, then figures of speech would be excellent FCAs. The teacher will create assignments that involve figures of speech in sensory-descriptive writing.

WT: Do teachers share all this information with students?

COLLINS: Absolutely! That is a critical piece of the process. I tell my teachers that their students should be able to come to them and say, "I think I am 'cooking' here" or "I still don't get transitions." The students should be able to recognize their problems and strengths. If students don't know how they are doing, there is something wrong. If they know what they are looking for, then they should know how they are doing in their writing. When you sum it up, it is really pretty basic.

WT: That is why it is so attractive. When you sum it up, it isn't that complicated.

COLLINS: I try to leave the complications out. When I present workshops, there are really only big topics: oral reading, focus correcting, and using past papers to teach new skills. I emphasize placing papers on the overhead projector and grading them as a group so that students can learn. For example, if we are discussing transitions, I take a Type Three paper before it becomes a Type Four, and we grade it for transitions as a group. I think that is some of the best teaching I have ever done. I tell them we are calibrating the papers. If I tell students that there are twenty-five points for transitions, they may not know what I mean. After students have seen me correct three or four papers, they will understand. The students can see what I mean.

WT: How do you choose the papers you will use on the overhead projector?

COLLINS: I used to tell my kids that if they volunteered to have their paper on the overhead, then I would fix it for them. If they volunteered, they were guaranteed a good grade. I had many mediocre and low kids who were "dying" to have their papers placed up in front of the class. I created a classroom atmosphere that said, "Let's stick to comments about the FCAs." The paper had not been evaluated by anyone else. As soon as the paper goes up on the overhead, the writer will see errors that were not seen before.

I also say, "Let's be gentle. If I see someone enjoying the criticism too much, that person's paper will be the next one up on the projector."

WT: What kind of doubts do teachers express about this portfolio approach?

COLLINS: They are often worried about not correcting every error. Spelling always jumps right out. What I come back to in this case is the oral reading. If teachers create the correct atmosphere, then when kids are reading their papers to one another, they can place a check in the margin when they see a spelling error. Then they can tell the writer about it when they discuss the paper. Some teachers fear that the kids will let everything else go and only pay attention to the FCAs. They may pay attention to transition words, for example, but forget about paragraphing, possessives, and all the other things that teachers want students to know. The fear is that if it is not graded, then students will not pay attention to it.

I come back to the fact that no kid likes to look "stupid" in front of his friends. Also, if kids think some of these papers might be up on the overhead projector, they don't want to look "stupid" in public. In addition, a teacher can always hand a paper back to a kid and say, "Hey, this isn't your best work." If the student answers, "Well, I had all three focus correction areas," the teacher's answer is, "Wait a minute! That doesn't mean you can misspell every other word. You can do better than that, and I'm not accepting this."

Once the teachers get involved in this kind of portfolio program, they discover that the students really love it. The most common feedback that I receive is "My kids really enjoy writing this year because they think they have a chance to be successful." The students feel more in control of their writing. It also allows them to take some more risks because they know they are not going to be "nailed" for them. Because they are enjoying it, their true voices emerge, more of their personality emerges, and their writing usually seems to be longer and more fluent. I have more and more teachers saying that they have seen no loss in the other skills, but teachers have to see it to believe it.

WT: Can you shed just a little more light on how teachers go about selecting the FCAs to emphasize during a school year? How do they strike a balance between the mechanical, style, and content corrections?

COLLINS: I suggest that teachers aim for a mix between organization, style, content, and mechanics. A teacher cannot cover all the FCAs in a school year. If a teacher covers eight FCAs in a year, that teacher has done a terrific job.

WT: Do you have any final comments or suggestions for teachers who want to use writing portfolios with their students?

COLLINS: Although I am best known for emphasizing the idea of focus correction, I would stress that first kids must learn to read their papers, listen, [illegible] develop their reader's ear. They can only do this through sustained practice. A teacher cannot do it three times in a school year and expect the kids to do it next year. Secondly, when teachers do focus correction, they may feel like they are letting their kids down because they haven't told them all their mistakes. If teachers can understand that by not grading a paper for all its mistakes, it becomes the world's best worksheet for the future. It is instant transfer.

The teachers must love the program. It's like new math. The teachers hated it, the kids hated it, and it died. If we can help teachers design a program that is reasonable and practical, they'll motivate the kids.

For more information on Dr. Collins' workshops and publications, teachers may call The NETWORK, Inc., at 1-800-877-5400.

Reprinted with permission from Language Arts,
Volume 64, Number 3, March 1987.

An Alternative Assessment Program in Language Arts

Dennis Searle
Margaret Stevenson

Background

The emphasis on accountability in our neighboring country has resulted, in many cases, in mechanistic approaches that have threatened to narrow and limit language arts programs to those skills and subskills that are most easily measured. In Edmonton we were keeping an anxious eye on the progress of such programs as Chicago Mastery Reading, knowing that such an approach was doomed to failure, but hoping it would do so before it "spread like a stain," as Jimmy Britton says, across the border and into our city. (It did!)

These concerns were certainly uppermost in our minds when, in the fall of 1982, the language arts consulting staff of Edmonton Public Schools received word that we were to develop statements of outcomes in language arts that were generalizable across elementary, junior, and senior high schools. We were also responsible for developing written achievement tests based on those outcomes for students in years three, six, and nine which were to be piloted, revised, and ready to administer in June, 1983.

Outcome Statements

At a hastily called meeting of the language arts staff in this large school district (70,000 students) we discussed our resources and the process we would use. Our resources included the expertise and knowledge of the staff members, a professor on exchange from the University of Alberta, a very supportive statement of rationale, philosophy, and objectives from the provincial department of education and "A Language Working Paper," a document outlining our own position on the language and learning of children. The guidelines agreed upon for the development of the outcomes statements were that they must reflect the integrated nature of language, they must be applicable from grades one to twelve, and they must not be great in number. As well, they must be specific enough to form the basis of the achievement tests. A tall order, but not impossible. The following eight statements were decided upon:

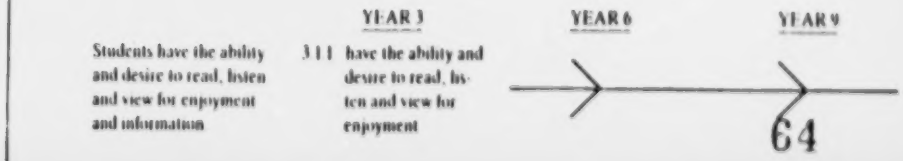
1. Students have the ability and desire to read, listen, and view for enjoyment and information.

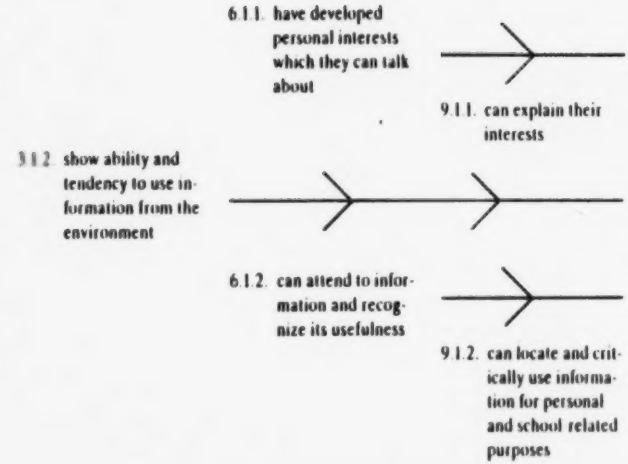
2. Students can and will read and compose literature to extend and clarify their own experiences.
3. Students can and will paraphrase, predict, and question ideas encountered in written and spoken language.
4. Students are aware of the literary tradition, its purpose, its complexity, its relationship to culture, and its value.
5. Students apply knowledge of written and spoken language in a variety of language situations.
6. Students use language to discuss language when constructing their own messages and when analyzing the messages of others.
7. Students express thought, feelings, and experience in written and spoken language.
8. Students organize and evaluate new experiences and ideas in written and spoken language.

These outcomes reflect our beliefs about language. Each one is stated to imply the active and purposeful use of language by learners. Literature, for example, is related to understanding both personal experiences and the universal experience of humankind over time. Knowledge about language is only considered useful if it helps students relate in a variety of situations, or helps them analyze and communicate effectively with others and understand fully and critically what others communicate to them.

Progress within Each Outcome

The next step was to develop indicators that would demonstrate student progress across years three, six, and nine as a basis for developing the tests. In creating these indicators, various related growth patterns were considered. The ability to use language effectively was seen as moving from the ability to focus on topics based on personal experiences, interests, and perspectives to include topics reflecting more public concerns based less on direct experience. Early skill using language in immediate contexts where the situation provided language cues would be augmented by the ability to use language in more remote contexts in which the students would have to be more aware of how the context made demands on language. Early use of implicit, chronologically ordered forms of language was seen to give way, on some occasions, to more explicit generalized statements of experience. An example of how the first outcome was seen to develop over the years is included.


YEAR 3
YEAR 6
YEAR 9
Students have the ability and desire to read, listen and view for enjoyment and information
3 1 1 have the ability and desire to read, listen and view for enjoyment


6.1.1. have developed personal interests which they can talk about
9.1.1. can explain their interests
3.1.2. show ability and tendency to use information from the environment
6.1.2. can attend to information and recognize its usefulness
9.1.2. can locate and critically use information for personal and school related purposes

It can be seen that students are expected to move from reading only for enjoyment to being able to reflect on their interests in reading. Similarly a critical element begins to develop in their ability to read for information. It is important to note, at the same time, that the continuing arrows indicate a recognition and desire that the early activity will also be evident throughout the following grades. The ability to talk about interests does not replace the basic activity of reading for pleasure, but augments it by allowing students to seek out reading they are apt to like, and to extend and share their pleasure through talk.

The Nature of the Tests

Our greatest challenge was trying to develop questions that would reflect, as much as is possible in a test situation, the language arts program that we supported in the schools. We wanted to be able to demonstrate what we expected teachers to do in their own classrooms. For example, we expected teachers consciously to assist students to develop language skills in all subject areas, to require students to express thoughts and feelings fluently and precisely in writing, to provide opportunities for personal response to literature, and for supported opinions. We expected teachers to ensure that their students had opportunities to write in different forms, for varied purposes and audiences. Our tests needed to reflect this kind of program, not only to support the teachers who were providing it for their students, but also because we knew the tests, if applicable, would affect program, since for good or bad this is the nature of testing.

It is important, then, that the tests call upon students to use language. At the same time we had to be aware of reasonable standards of use. We wanted students to write but realized we could only provide the time and resources nec-

essary for first draft writing. We also recognized that context is part of purposeful use and tried to provide a context of audience and background to assist the students. In the Grade Three test, for example, context was developed by integrating language arts and social studies questions.

The result of this integration is evident in these sample questions from the Grade Three test:

> Imagine that you have moved from Edmonton to another community. Your new home will be either on a farm OR in a 'special community.' Write about your first day in your new home, and how you felt about it.

This question related to the seventh outcome. Students were expected to adopt a viewpoint and feelings which would be probable for someone who had moved, and to express their viewpoint in personal writing. Not all writing was graded for convention, but this piece was. Students were expected to demonstrate ability to use the conventions of capitalization, punctuation, and spelling. Absolute correct usage was not as important as the recognition that conventions helped demonstrate the internal organization.

Another item, which grew out of a discussion of settlers working together was:

> List three rules to remember when you have to work with others and you want to do your best.

This item was scored for social studies on the basis of the awareness of what is necessary to work together. In language arts the scoring focused on the ability of students to use the format and language of rules.

The developing sophistication expected of students can be seen in the Grade Six test. The examples given are from a series of activities related to an extended writing task.

> The first day of school each year is quite an experience. Think back to a first day of school and in the space below, *list* some of the things you remember. You might include what you did or who you met. There are probably other details that you connect with a first day of school. Include these as well.

We thought that this was a memorable experience common to all students. The task was graded to assess outcomes five and eight. It was expected that students could use a brainstorming format which allowed them to list a wide range of ideas. More importantly, their ability to organize and evaluate the experience through the selection of significant details was checked. The activity was also included to prepare the students for the actual writing task, which was:

> Suppose you meet a child who will be beginning Grade One who asks you some questions about going to school. Write some information and advice that will help the child get off to a good start on the first day of school. Remember that the parents will read your advice to the child.

The task calls for writers to be conscious of their role as an "expert" giving advice, and of two audiences, the Grade One pupil and the parents. The wording and advice must match a Grade One student's comprehension and concerns, and the attitude must meet with parental approval. The ability to match language to a particular language situation is tested; at the same time the ability to express thought, feelings, and experience is assessed through the clarity of the advice given.

Students had been told that their work was to be a first draft only. However, to test the students' familiarity with the writing process, a final question was added:

> State two ways that you could improve what you have written in question 17 (the advice) if you had a chance to write a final copy.

Any two plausible means of improvement were accepted. There is diagnostic information available in this question as well, in that the level of concern can be recorded. If students see spelling checks or improved handwriting as the way to improve rather than improvement through added content or more effective style, then we know what message about writing is being received by students. We felt that this question would support teachers who used teacher or peer conferences and gave students opportunities to talk about work in progress.

It was necessary to develop a common concern and to provide something to think about in order to create a test for Grade Nine. Consequently, the Grade Nine test focused on a theme, Television. One source of input was an essay about television. Students were asked to react as follows:

> Summarize in your own words, the opinion that the author has about TV and about people who watch it.
>
> Use a straight, solid line to underline a section in the essay that is particularly effective. Explain why you consider this section to be effective.

The first question related to the first and third outcomes. Students were required to demonstrate both ability to get information and to paraphrase. The second question required students to demonstrate an awareness of what makes writing effective. Students who could talk about writing, naming qualities of writing or parts of the essay, gained additional marks.

In response to a poem about television, students were presented with this question:

> Literature often reminds people of experiences they have had or stories about someone else's experience. Describe an experience you are reminded of by the poem.

This question is related to the second outcome and asks students to use literature to help see their own lives. The experience of empathizing with literary characters and themes will help them to identify and relate a personal experi-

ence with enough detail to allow a reader to understand the experience and its relationship with the poem.

Results

It must be stressed that these tests were designed to evaluate a school district's success, not the success of individual students. The aim in the scoring is to focus on the eight outcomes rather than record all the individual strengths demonstrated in each answer. At this stage we can say that the tests do measure and demonstrate growth. Although the tests were designed for grades three, six, and nine, the items have been administered to students from grades two to ten. A quite definite pattern of growth has emerged.

One comfort has been that the first outcome relating to reading for pleasure and information has consistently scored highly. The tests also revealed that, as in most school districts in North America, the attention given to actual extended writing has been limited. Originally, extended writing tasks were not handled as competently as hoped. Increasingly, however, there is more attention being paid to writing in the classroom, and this attention is beginning to be reflected in the evaluation. Processes such as learning logs and brainstorming are being encouraged as teachers see these demonstrated in the tests.

Schools which have been moving towards the provincial goals and using "A Language Working Paper" as a basis for their programs are encouraged by seeing a testing system which supports their directions.

A further benefit has resulted from the approach taken to marking. Because writing is included, the tests cannot be machine scored; instead teachers have been employed at the beginning of the summer to work in a three- to four-day marking session. The experience of marking has proved to be a very effective language arts in-service in the district. Teachers are trained to score extended writing holistically. They discuss the attributes of shorter responses knowledgeably as they interpret and use the scoring guide. Reading papers written by many students of varying competence and ability helps teachers understand where their own teaching and students fit into the total picture.

In addition, approaches to evaluation taken in the tests provide models for a broader approach to evaluation in the classroom. Knowledge of the outcomes and their implications at each grade level is developed.

Impact on Language Arts Instruction

There is no doubt that the achievement tests are having a positive influence on language arts programs; the decision to require students to express their thoughts, feelings, and responses to literature totally in writing has sent a message to the schools. Many school staffs have made writing a priority which is reflected in increased attendance at district writing workshops and requests for

long-term assistance. In-service material has now been completed that provides a general introduction to each outcome from K–12 and provides specific examples for the major developments. This material will be provided to school staffs along with suggestions for its use in observing and recording students' language learning. Schools are already moving, on their own, to identifying growth in writing over the school years.

Final Comments

It would be foolish to assume that these tests are, or ever could be, perfect. The tests and the marking guides are monitored and improved yearly. In addition, it is stressed that teachers and administrators must always be aware that tests provide only one piece of information. Their own observations, anecdotal records, conference notes, check lists, and records provide a valuable source that must always be taken into consideration. In fact, our next project is to gather non-paper-and-pencil assessment ideas and make them available to all district teachers. In this way, other aspects of the program can be evaluated more effectively. Tests such as these do not provide reliable information at an individual level and were not intended to. They do allow schools to identify strengths and weaknesses in program and to monitor improvement.

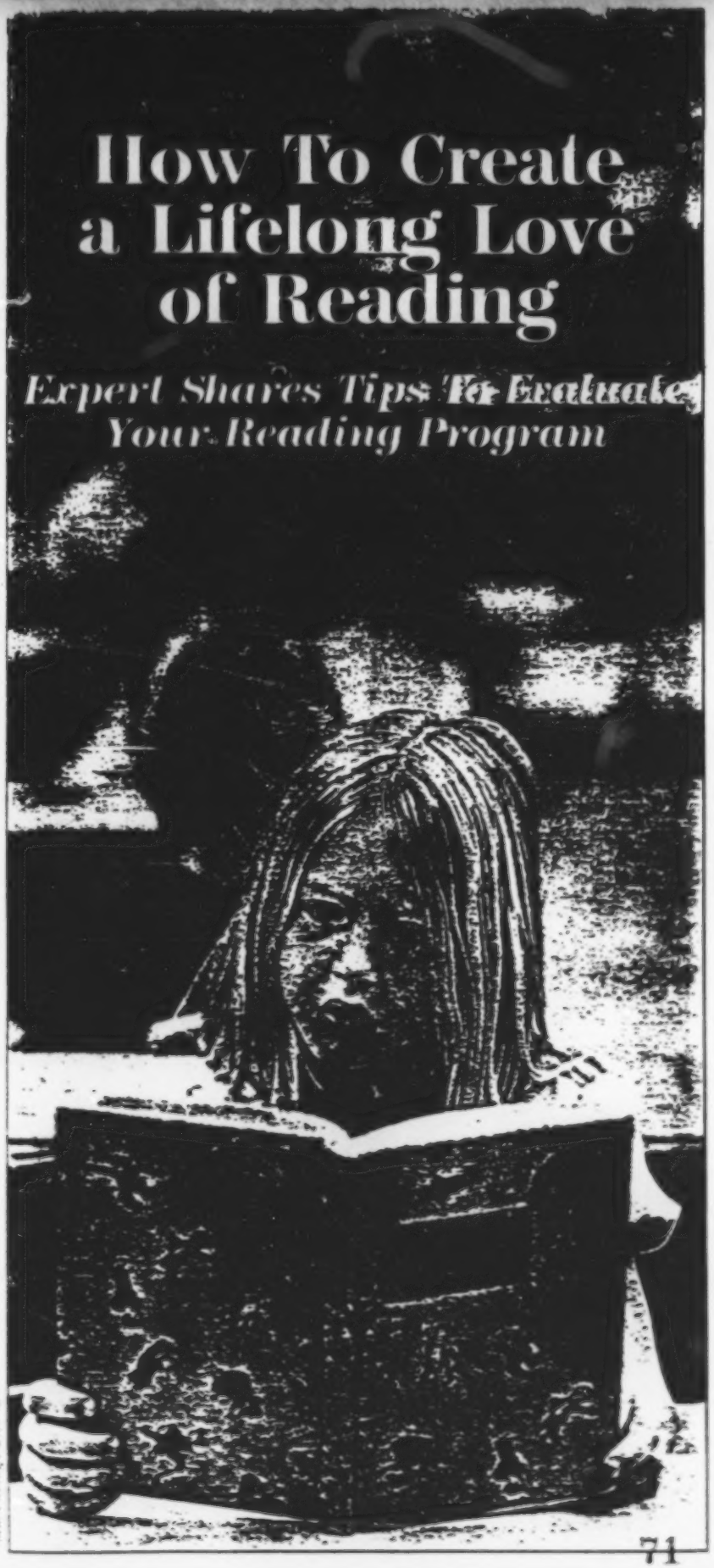
What the tests have done is demonstrate a unity among beliefs about language, program objectives, instructional procedures, and evaluation. In fact, the cooperation between evaluation services and subject area specialists is fundamental to the success of these tests. District-wide achievement testing may not be the only way to demonstrate growth and evaluate a school district's performance. If, however, such an approach is adopted, we feel it is imperative that people working within the district become involved to make the evaluation work for the benefit of the students and teachers.

Dennis Searle teaches at York University in Toronto. Margaret Stevenson is a language arts consultant for the Edmonton Public School Board, Alberta.

Reprinted with permission from The School Administrator, May 1992.

How To Create a Lifelong Love of Reading

Expert Shares Tips To Evaluate Your Reading Program



BY SIDNEY J. RAUCH
Professor Emeritus of Reading and Education, Hofstra University

How do you know if your school district's reading program is a good one? Do you judge it by the results of standardized test scores, classroom observations, assessment of children's work (such as portfolios), comments of teachers, pupils, and parents—or all of the above?

The International Reading Association endorses the latter approach, stating in a 1988 board of directors report: "Reading assessment must reflect recent advances in the understanding of the reading process. ... [I]nstructional decisions are too often made from assessments that define reading as a sequence of discrete skills that students must master to become readers. Such assessments foster inappropriate instruction."

In the view of Donald Graves, an authority on children's writing, educators fail to ask big enough questions of reading studies. "We need to determine only one thing: How do we help create a lifelong love of reading?" he asked during a speech to the New York State Reading Association.

As school leaders evaluate reading programs and collect and analyze data, they should not lose sight of that fundamental objective.

Reading programs in elementary schools need constant evaluation. This should compel administrators, teachers, and parents to take a comprehensive, unbiased, and cooperative look at their program and decide what modifications are needed. The effort to determine what is "valuable" or "of worth"—a concept developed by B.R. Worthen and J.R. Sanders, co-authors of *Educational Evaluation: Theory and Practice*—typically results in better reading programs.

In general, five major steps are involved in the evaluation of reading programs. They are:

- Clarification of the roles of evaluators,
- Collection of data,
- Analysis of data,
- Reporting of data, and
- Follow up.

Sam Ludwig © 1992

How Does Your Reading Program Stack Against this Checklist?

Administrators and instructional supervisors who evaluate school reading programs could use this checklist for guidance.

The checklist was prepared by Sidney Rauch, professor emeritus of reading and education at Hofstra University and winner of two Children's Choice Awards in 1991 from the New York State Reading Association for his works of fiction, *The Visitor from Outer Space* and *The Return of B.B.*

Motivation

- Has the teacher created a desire and/or purpose for reading?
- Does the teacher supply needed background information and clarify key vocabulary and concepts?
- Does the teacher use the prior knowledge of children to improve comprehension?

Clarification

- Does the teacher encourage children to predict or anticipate the lesson's context?
- Does the teacher attempt to "build bridges between the new and the known" (a concept introduced by P.D. Pearson and D.D. Johnson, authors of *Teaching Reading Comprehension*)?
- Does the questioning go beyond factual information into the area of critical thinking and evaluation?
- Does the teacher recognize and attempt to meet the special needs of individual students?

Application

- Is there an attempt to apply strategies and information gained to other content areas?
- Are youngsters made to see the relevance of the material covered to everyday life situations?

Satisfaction

- Is there concrete evidence that the students have learned something?
- Does the students' behavior indicate feelings of satisfaction and understanding of the information gained?
- Does the observer sense what A.H. Maslow, author of *Motivation and Personality*, refers to as "the emotional aspects of cognition, the lift that comes with insight, the calming effect of understanding?"
- Does the lesson advance our primary objective of a reading program, i.e., "to create a lifelong love for reading?"

Clarifying Roles

Many school administrators wonder, "Who should be responsible for the evaluation? Should it be done in-house or by a team of outside evaluators?"

Those who favor the in-house approach frequently cite such factors as familiarity with the program and personnel involved, the less threatening atmosphere since the evaluators are known, and the evaluators have a better understanding of the situation realities. The recommendations that follow a thorough evaluation must take a realistic view of not only what should be done, but what can be done, in light of budget limitations.

Those who favor the use of outside authorities cite the expertise and proven experience of these individuals as evaluators and the unbiased approach they bring to the task.

Whatever the choice, your evaluators should ensure the following:

- Objectives and procedures must be clarified before the evaluation begins. Teachers must understand the need for evaluation. Their cooperation and support are essential. A successful evaluation involves full faculty participation and self-study.
- Teachers must understand that evaluation focuses upon the reading program, not the strength and weaknesses of individual teachers. The purpose is neither to rate and compare teachers nor to rate and compare the performance of individual schools.

The following excerpts, drawn from a letter addressed to the teaching staff at one school, might be helpful in allaying anxieties about evaluations: "This study will NOT be a teacher evaluation project, nor will it in any way rate school against school. The study staff won't come with panaceas nor with bias toward any special 'method' for teaching reading. It will not be an 'efficiency' study, nor will there be effort to promote uniformity of instruction in every classroom.

"Since this study must rely upon the cooperation and involvement of all personnel, it shall be conducted so that each teacher's ideas, questions, and appraisal of needs will at any time be welcomed. Don't hesitate to stop consultants in the halls or invite them to meet informally with your grade group. They will welcome the opportunity. We shall keep all teachers and administrators informed as the study develops, and will welcome your suggestions as well as your questions."

Collecting Data

The data collected will involve all or most of the following: standardized test scores, observation of classroom performance, assessment of children's work, interviews with administrators, teachers, parents, librarians, reading and learning disability specialists, school psychologists, speech therapists, and others.

In addition, interviews with pupils from grade 3 on can provide helpful data. Opinion surveys provide an opportunity for self-study and can identify the strengths and weaknesses of the reading programs.

Standardized Tests

Though much criticism has been leveled against standardized tests, they remain a useful and valid source of objective measurement. School boards and parent groups continue to demand "hard data," despite the values of informal assessment.

Standardized tests can be a valuable source of information if you follow these caveats:

- Standardized test scores are not absolute values, and should never be considered as though they were. At best, they are estimates of a student's reading ability.
- No youngster's educational future should be based on the results of any single test score.
- Standardized test scores are one form of data to be placed alongside other data in evaluating a program's strengths and weaknesses. Generally, the more data obtained, the more accurate the interpretation.

Classroom Observation

Of all the data collected during an evaluation, this is the most significant. Classroom observations are the heart of reading evaluation. A representative number of teachers should be observed and their performance evaluated. To evaluate without observing teacher performance in conjunction with pupil response and behavior is equivalent to reviewing a book without reading it or a play without seeing it.

Each observation should be followed by an informal meeting with the teacher observed, so that person has the opportunity to comment about the purposes and procedures of the lesson. For guidance, evaluators might use a short checklist built around four key qualities: motivation, clarification, application, and satisfaction. (See related story, page 28.)

Using Interviews

If a feeling of trust is established between the evaluator and the interviewee, valuable information can be obtained about the strengths and weaknesses of the reading program. The following questions can elicit helpful replies:

- "What can we do to help with the program?"
- "What reading strategies or techniques do you find most helpful as you work with your children?"
- "Are there any materials that you need?"
- "Put yourself in the place of the evaluator: What recommendations would you make to improve reading instruction?"

Pupil Assessment

It is important to evaluate students' abilities to use and interpret language in realistic situations. Thus, the assessment of children's reading performance and writing samples accumulated during the school year can provide valuable evidence of students' learning.

Pupils' self-evaluation of their own work is an integral part of this process.

Using Surveys

A successful evaluation must involve self-study. The distribution of one- or two-page opinion surveys to administrators, teachers, parents, and pupils can help you obtain insights into the program. Recipients should be briefed as to the purpose and contents of the survey. All involved should be reminded their input is crucial to the success of the evaluation.

Analyzing Data

Evaluators are in constant contact with each other during the evaluation. Observations and impressions are shared, keeping in mind a primary objective of the study: What can be done to improve the reading program?

Following the evaluation, each evaluator prepares his or her own report based on classroom observations, personal interviews, and analysis of data obtained from opinion surveys. A measurement specialist on the evaluation team should analyze standardized test scores from a five-year period. Dramatic changes can occur from factors other than an individual's improvement or decline, such as shifts in school population, use of new tests, particularly one with a radically different format, or an inordinate amount of time spent on test preparation.

The study leader has the final responsibility for analyzing the reports of individual team members, collating and organizing the data and suggestions for improvement, and preparing the final report. The report should be reviewed by team members, then submitted to the school superintendent.

Reporting Data

When you analyze the data, provide both an oral and written report of the findings of the evaluation to the administrators' council, teacher representatives, and the school board. The board, if it wishes, can hold an open meeting for all parents where questions can be asked of the evaluators. Copies of the report should be made available to all teachers. Several copies can be placed in the teachers' professional library or a special section of the school library.

Vital Follow Up

Too often, evaluations (particularly by outside evaluators) prove to be a hit-and-run routine. The evaluators do their job and never are heard from again. A scheduled follow up can take care of this weakness.

A schedule of follow-up meetings can enhance the effectiveness of an evaluation. School personnel can be informed at the beginning that approximately one year after the evaluators report to the school board, there will be follow-up meetings with committees (administrators, teachers, parents) to determine what progress, if any, has been made.

One of the best ways to do this is to compare the list of recommendations against the efforts that have been made to follow the recommendations. Since the selected committee will have been informed in advance of the follow-up meetings, they will have ample time to collect data and prepare their report for the evaluators.

In most instances, evaluation has a positive effect on the reading program. It compels administrators and teachers to take a closer look at their philosophy, their methods and materials, and their children. This comprehensive examination generally results in a stronger program.

BEST COPY AVAILABLE

Reprinted with permission from The School Administrator, June 1989.

Can Cooperative Learning Be Evaluated?

Yes, If the Observer Knows the Keys to Its Success

BY DAN WATSON AND LYLE RANGEL

A popular teaching strategy of the 1980s, cooperative learning, may continue well into the 1990s for three reasons:

- It improves academic performance for all students including minorities;
- It enhances ethnic interactions and reduces racial frictions; and
- It increases English language proficiency for English-as-a-Second-Language students.

Photo courtesy of Johns Hopkins University

Due to cooperative learning's popularity, site administrators now formally observe teachers who use some form of collaborative or cooperative learning in their lessons. Observing and evaluating a cooperative classroom, however, can challenge the site administrator. For one thing, at least nine different cooperative learning models are practiced today.

The big question administrators face is can teachers using cooperative learning techniques actually be evaluated?

According to Spencer Kagan, a professor at the University of California, Riverside, and one of the better-known advocates of cooperative learning, "Teachers using cooperative learning cannot be evaluated."

Kagan believes a teacher using cooperative techniques can be coached, but should not be evaluated.

We believe, however, despite the numerous models and Kagan's misgivings, evaluating teachers' use of cooperative learning can be done and done well if the evaluator knows what to look for.

Identify the Basics

Successful evaluation depends on the evaluator knowing the essential elements of the teaching strategy and comparing the known model with observation.

In cooperative learning, the most frequently used models have five common components: heterogeneous teams; teammate interdependence; interdependence of teams; accountability; and activity appropriateness. The administrator observing cooperative learning would be looking for these characteristics. Each is described briefly below.

- *Heterogeneous teams.* Teams should be composed of four to six members and represent a true cross section of the classroom. There should not be "high flyers" and "low flyers" teams. Instead, teams should be balanced by race, gender, and academic achievement.

In the classroom, the seating arrangement should encourage team members to interact. Teams should sit at common tables or with seats together, not in straight rows. Team identities or nicknames should be readily apparent.

- *Teammate interdependence.* Interdependence of teammates means

that the success of each teammate depends on the success of every member, that is, they sink or swim together.

When interdependence exists, students should be aware that they play a unique role on the team and that the team cannot succeed without their efforts. In almost any sport, each team member can identify their position. Students in a cooperative learning setting should be able to do the same. Team members should be able to state, "I'm the recorder," or "I can't do my step until Maria does hers."

Teammate interdependence can also be observed in student behaviors, such as working closely together; discussing what they have to do; completing drills and practices to increase accuracy; helping one another with the answers to questions; using materials together; and, most importantly, assisting each other to learn.

● *Team interdependence.* Not to be confused with interdependence of teammates, the various teams in the classroom should not be locked in a struggle to overcome their enemies—their classmates on other teams.

Through cooperative learning, team members learn while helping the team attain a simple goal—the academic success of its members. The goal is not to seek and destroy the competition.

As we've all seen, when kids compete in some sports activities, intense team competition can generate strong negative actions against the opposing team.

In the cooperative learning classroom, such competition can be detrimental. Rewards should be set up so the whole class is rewarded for combined achievements after an extended period of time. (A good example of this kind of reward is a thermometer chart used to gauge progress toward a particular goal.)

Students should also be aware that their current teams are only temporary. Teams should be rearranged at a maximum of every six weeks to allow extensive interaction among classmates and to avoid unexpected "unfair" teams from dominating.

● *Accountability.* One of the most talked about issues in cooperative learning is accountability of team members. Educators continually worry that one team member is going to do all work and the rest will ride on that person's coattails.

The classroom evaluator looks for two types of accountability. First, is the team rewarded for achieving a group goal? Rewards could include: single grade for the group's efforts; bonus points added to an individual's score when the group attains a specific criterion level; non-academic rewards for attaining a group goal; and team recognition for attaining a goal.

Second, is each team member held accountable? Each student is expected to improve in the task. Examples are: improvements on a criterion-referenced test from pretest to post-test; and grades given to individuals based on their score compared to individuals of the same "ability" level (rather than grade based on a whole class curve).

If students were observed riding on the coattails of a single person on many teams, a post-observation conference between the teacher and evaluator should be scheduled to discuss this problem.

A teacher allowing this to happen in the classroom might need more intensive coaching or directing to use specific cooperative techniques, such as Student Teams Achievement Divisions, Teams-Games-Tournaments, Jigsaw I or II, or Groups of Four. These structures are more specific in their organization and have proven successful in handling teams that turned into solo acts.

● *Activity appropriateness.* Though cooperative techniques have been used with almost every type of content lesson, not every task can be completed cooperatively.

For example, while cooperative techniques can be used to help stu-

Photo by C.W. Dunn

Photo by Susie Fitzhugh © 1989

"...the success of each teammate depends on the success of every team member...."

dents become more accurate and quicker with math facts, using cooperative techniques during a timed math facts test probably wouldn't work.

Indications of inappropriate activities are: students spend the majority of time avoiding the task; students are unclear on what they are to do; students do not appear to have enough information to complete the activity; or students do not have the social maturity to deal with the complexity of the task they are assigned.

Photo by C.W. Dean

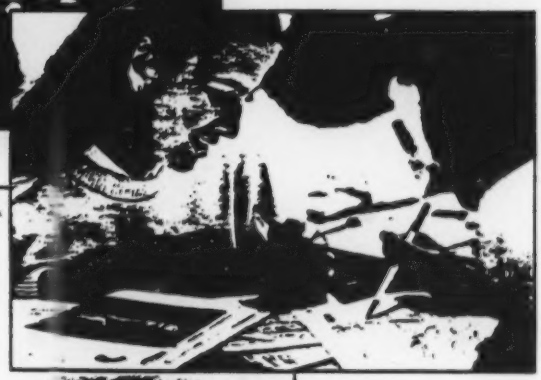


Photo by Susie Fitzhugh © 1986

Photo by Susie Fitzhugh © 1989

"...the amount of time the teacher is instructing should not dominate the class."

Other Essentials

Cooperative learning cannot replace direct teacher instruction in most educational settings. Teachers using cooperative approaches also need to instruct.

However, in cooperative learning, the amount of time the teacher is instructing should not dominate the class. The teacher needs to motivate learners, state the goals of the instruction, and teach the content.

The teacher should use guided practice to check for basic mastery of content and ask students to complete independent practice through cooperative learning groups.

Cooperative learning requires a definite shift in the role the teacher plays. During the instructional period, the teacher is the focal point and information flows from teacher to students. During the cooperative component, however, the teacher is a facilitator of students. Moving from group to group, the teacher acts as a resource to the teams. At the close of instruction, the teacher again dominates, seeking closure for the class.

Using cooperative learning is an excellent sign that teachers are keeping up with current trends in education. Without support and evaluation, however, a teacher may be implementing the technique incorrectly or failing to maximize its use.

Dan Watson is the coordinator of Student Well Being, San Diego County Office of Education, San Diego, California. Lyle Rangel is the principal at Wegeforth Elementary School, San Diego City Unified, San Diego.

Resources

Burns, Marilyn. "Groups of Four: Solving the Management Problem," *Learning* 9:46-50, 1981.

Cohen, Elizabeth. *Designing Groupwork: Strategies for Heterogeneous Classrooms.* New York: Teachers College Press, 1987.

Dishon, Dee and O'Leary, Pat W. *Guidebook for Cooperative Learning: A Technique for Creating More Effective Schools.* Holmes Beach, Florida: Learning Publications, 1984.

Graves, Nancy and Graves, Ted. *Cooperative Learning: A Resource Guide.* Santa Cruz, California: International Association for the Study of Cooperation in Education, 1987.

Johnson, David W. and Johnson, Roger T. *Learning Together and Alone: Cooperative, Competitive, and Individualistic Learning* Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1987.

Kagan, Spencer. *Cooperative Learning Resources for Teachers.* Mission Viejo, California: Resources for Teachers, 1987.

Slavin, Robert E. *Using Student Team Learning.* Baltimore, Maryland: The Johns Hopkins University, 1986.

BEST COPY AVAILABLE

Reprinted with permission from the Spring 1991 issue of the AMERICAN EDUCATOR, the quarterly journal of the American Federation of Teachers.

BEGINNING RESEARCHERS

How To Make Report Writing More than Copying from World Book

This article illustrates the intersection of teaching, learning and assessment through a description of experiences designed to build students' skills as researchers.

BY DONNA MAXIM

"MY STUDENTS are so wonderful," Nancy Nash remarked when I stopped by her classroom after school one day. "They've just completed their research projects, and I've learned so much."

"What do you mean?" I asked.

"I remember doing research projects last year, during my student teaching. What a disaster! Now, I can see why some things didn't work. First, I'd chosen the topics. Kids had to write about a U.S. state—the focus of the research had nothing to do with what we had been studying. Then they got out the encyclopedias and started to copy. But this year my students selected their own topics, and their study evolved naturally from our science curriculum. When they went to the library, they didn't rush to the encyclopedias. They looked for resources on the shelves and used the card catalogue subject drawers. It was really exciting. They took notes without copying. And then they . . . ."

I smiled as I listened to Nancy talk about her fourth-grade researchers. I knew the work they had done as third graders: all the literature they had read and listened to, all the carefully planned activities that had helped them to succeed in fourth grade. I smiled because it's not often that a third-grade teacher hears words of praise for her former students from an upper-grade colleague and because I knew that the beginning research techniques that I taught had helped my kids become confident, skilled researchers.

Third graders enter this metamorphic year as hands-on investigators. Over many months they move toward independent research by using reading and writing as vehicles to make new discoveries about themselves and the world around them. But before my third graders ever become involved in independent research projects, I model a variety of procedures used by researchers. I demonstrate, students practice, and we share our discoveries as we investigate new topics across the curriculum. I show them note-taking and interviewing strategies and teach them how to use log entries to predict and then evaluate information brought to the classroom by visiting speakers. I also help them set up individual research notebooks, each complete with a resource list, webs, questions to pursue, and notes. Then they plan final products—not full-blown research papers but manageable formats that beginning researchers can control to display their new-found information. Finally, students discuss what helped them during their research and the discoveries they made during the process.

This article is organized around the range of research activities that I sponsored to expose my students to a vari-

---

Donna Maxim teaches at the Center for Teaching and Learning in Edgecombe, Maine. This article and the sidebar by Nancie Atwell are reprinted with permission from Coming to Know: Writing to Learn in the Intermediate Grades, *edited by Nancie Atwell and published by Heinemann Educational Books, Inc., 1990. The book may be ordered from Heinemann by pre-paying $16 plus $3.50 shipping and handling (361 Hanover St., Portsmouth, NH 03801-3959. Order #08500. Telephone 1-800-541-2086).*

ety of research principles and procedures and to give them practice in note taking, observing, interviewing, and questioning. I will also describe project formats that third graders have used to share their new information.

EARLY IN the fall, I introduced methods for note taking that helped my students avoid both plagiarism and an overreliance on the encyclopedia as a source of information. I began an ocean unit in science by reading aloud the author blurb from the back cover of Vincent Dethier's *Newberry: Life and Times of a Maine Clam* and asking why the information about Dethier's background is important enough to appear on the book jacket. Dethier is a professor of zoology, a person who studies animal life, and although *Newberry* is a fictional story, within it are embedded many facts that he learned through his own research.

During the first read-aloud, the class listed on chart paper the sea creatures they had met and the facts they had learned about each. The following day, after I had read another chapter, I asked children to write down the facts they had heard as individual log entries. During each of the subsequent daily readings, students did the same. After each log entry I also encouraged them to list questions that the story made them wonder about, questions that they might need other resources to answer. In this typical log entry, Dixie has written six facts that she recalled from the story and added two questions that she wondered about.

10/27 8:55

1. Starfish have suction cups.
2. Clams have two shells.
3. Starfish can take their stomachs out of their body to get food.
4. Starfish eat clams.
5. Clams don't like starfish.
6. Starfish can pull clams and mussels shells apart.

How do starfish take out their stomachs?
What do starfish eat besides clams and mussels?

On another day, Steven wrote:

A clam can build his shell bigger on the edges. A shrimp can shed its shell when it gets too small. A shrimp can eat a mussell.

Based on Steven's log entry, we discussed the fact that a shrimp is a filter feeder and, in the story, could not have eaten the mussel on Newberry's muffler. Over several readings we discussed fact versus fiction as we posed questions: Are shrimp filter feeders? Do shrimp really eat mussels? The students searched out other sources of information to answer their questions, and we concluded the reading of *Newberry* with a "Clam Facts" bulletin board.

Another book I read to students during science class was Arthur Myers's *Sea Creatures Do Amazing Things*, a collection of short articles about the strange habits of ocean animals. In a three-minute log entry, Sarah recorded the facts she had learned about a fiddler crab.

11/6 11:10

1. Fiddler crabs do burrow in sand.
2. A fiddler crab's biggest claw does look like a violin.
3. A fiddler crab's small claw does look like a bow.

In another note-taking session, children listened to a Random House tape and filmstrip on Newberry Medalist Margaret Henry. In her own words, Heather noted what she regarded as the important facts about Margaret Henry's life and the thoroughness of the research that goes into her books for children.

11/23 9:30

Margaret Henry is a Newberry Award winner. She loves horses. She wrote lots of books. She hated doing maps. She wrote mostly animal stories. She likes animals a lot. She traveled a lot. She likes children a lot. She went to horse ranges. She went to the Grand Canyon. She went to Italy three times in a year. She gets a lot of letters. When she was ten, she sold a piece for twelve dollars.

In each of these log entries, students took notes without being tempted to copy: to read and write at the same time. I gave them practice in listening to build up their skill and confidence as note takers and to build up to taking notes from print resources. Then I turned to *Zoobooks*, a monthly magazine published by Wildlife Education Ltd., and they had their first experience taking notes from their own reading.

Each student chose one back issue of *Zoobooks* but was not permitted to open it. I asked them to look at the front cover and generate questions they thought they might find answers to in the magazine. Andrea's questions about the wolf set the stage for her active participation as a reader of her particular *Zoobooks*.

12/6 2:00

1. What do wolves eat?
2. How do they live? (What is their house made of?)
3. How many years old can it grow up to?
4. Can they eat anything?
5. How many miles per hour do they think they can run?
6. Can they climb trees?
7. Would they eat another wolf?
8. Would they eat a dog that looks like them?

After they recorded their questions, I sent the students away from their desks, their lists of questions, and their pencils, to read for ten minutes. When the time was up they returned to their desks, this time leaving the *Zoobooks* behind. In their logs they wrote any answers they had found to their questions and added any new information they recalled. Students were beginning to take notes without copying. For many, this was difficult; they wanted to retrieve the magazines and copy the relevant information, something that I did not allow. A couple more sessions of writing without reading showed my third graders that they could read and then take their notes.

Through such activities as these, I teach my students how to listen to, read for, and respond to information. They learn how to recognize important facts, use more than one resource to verify information, take meaningful notes, and record facts quickly and easily. Instead of copying someone else's words, they are beginning to see themselves as researchers who take notes. The chances are good that they will not become victims of the plagiarism syndrome.

AT THE same time that the children practiced note taking, we continued our ocean study with slide shows, a visit from a speaker from the Department of Marine Resources (complete with live specimens), a field trip to Reid State Park, charts listing ocean life forms, and numerous read-alouds and discussions of fiction and nonfiction books as well as poetry. Eventually, the students

were prepared to choose a sea creature to research and report on.

The students began their independent research by using resources from both our classroom and the school library. Each student set up a research notebook, a three-ring binder that allowed easy access to information. The first page was a title page; the second page was a list of questions about their creature; and the third was reserved for listing resources by title, author, and page number. Students recorded notes on the remaining four or five pages. It was a simple, specific format for collecting, organizing, and keeping track of new information.

During the research process, Alice Fossett, our school librarian, helped students become familiar with resources within the library. She planned activities to map library resources and to use the card catalogue as well as the *National Geographic* index to locate magazine articles on chosen topics. Students kept running notes in their three-ring binders.

1/11 1:50

NOTES

Bony meat.
Large eye.
14 inches long.
6-10 inches
1/2-1 pound.
Bait on bluefish.
Sweet tasting.
Schooling fish.
Found Penobscot River.
60,000-100,000 eggs.
Goes to fresh water to spawn.
Lives in salt water.
1 ft.
Lobster bait.
Grayish-green.
3 years.
Silver belly.
Related shad.

Gearry's log entry shows how he used the ten minutes that I gave my students, after reading in the library, to record information that he recalled from his reading. His notes had changed since the beginning of the year. Now he used his time to capture many brief pieces of specific information, rather than wasting time composing long sentences. He was also selective, deciding which facts were worth keeping. In a mini-lesson, Gearry discussed his notes with the rest of the class via an overhead transparency, explaining why he wrote only concise phrases and how this technique helped him to get more information down.

After two weeks of researching and sharing their notes, the students wrote up their facts in the form of acrostic poems. I had read about a high school teacher's use of acrostics, or biopoems (Johnston 1985), and felt that this would be a simpler genre for third graders to control than the complex format of a traditional research report. An acrostic uses the letters of the word that is its subject as the beginning letter of each line. I added one rule: Each line must be a fact about the poet's creature.

LOBSTERS

Lays up to 5,000 eggs
One claw is to crush their prey
Bait is what they use to catch lobster with
Sometimes if you hold one it might pinch
Turns red when it's cooked
Eats pogies, mackerel, bluefish, and redfish
Return females with eggs to the ocean
Scavengers!

Betsy wrote lots of facts in her acrostic and showed what she had learned. Wendy, a classmate, listened to another student's draft of his acrostic in a group share meeting and declared that one line—"Clams are neat"—was an opinion, not a fact. From that day on I had many keen observers of the differences between facts and opinions.

I also taught the children interviewing techniques. I began by modeling an interview in a mini-lesson. I had prepared a list of questions about a marine creature that one of my students was studying, and I interviewed her about it, with her permission, in front of the class. Then the other students, working in pairs, wrote interview questions about their marine animals as log entries and practiced interviewing each other. Afterward, volunteers shared reactions to their interviews with the class, and we noted their strengths and weaknesses on a chart.

My students decided that good interviews featured lots of questions, were prepared ahead of time, and included an interviewer's explanation of why the question was being asked. Then I arranged for them to interview fifth graders who were also studying marine creatures. Because the older kids' topics were also self-selected, the range was so great that we had few matching topics, so each student chose a fifth grader's topic and prepared a good list of marine-related questions. The interviews worked well, showing my nervous students that they could question older children; they also served to generate questions that could be used in their own research. (Some third graders discovered that they knew as much as some fifth graders about a particular subject.) A drawback was that the third graders had no personal need for the information they were gathering.

Interviewing techniques flowed over into social studies class, where I was reading aloud *Marching to Freedom: The Story of Martin Luther King, Jr.* by Joyce Milton. The children had already discussed their dreams, including dreams for the Boothbay school, so I asked them to brainstorm questions they might ask parents, teachers, school board members, our principal, or some other adult about their dreams for schools. Each student wrote a log entry setting up an interview. Then, as a class, they discussed the discoveries they had made: how they needed to explain who they were and why they were calling when approaching people they didn't know; how a good friend was approached differently from a stranger.

BEST COPY AVAILABLE

Again, they role played interviews in log entries, and we charted whom they might contact. We also charted how they might contact particular individuals and listed questions they might ask. Students generated their own final guidelines for interviews and listed these on a separate chart:

Tell or Ask:

- Your name
- Where you're from
- Your grade
- What the project is
- Questions to ask
- Follow-up questions
- Thank you

The interviewing techniques I modeled and discussed and the actual interviews they conducted gave my third graders valuable insights into another avenue for research. They needed to look to primary sources—to people—as well as to good literature. The various interview activities helped them overcome their nervousness and take more control of their search for information.

ANOTHER RESEARCH practice I initiated early in the year was keeping logs of observations. Such logs show students how researchers set up and record information, help them to discover the differences between observations and predictions, and allow them to see themselves as resources in their own research.

In the fall, I asked each student to choose a tree in our playground to observe throughout the four seasons. The students recorded observations of their adopted trees and the surrounding environment at the beginning of each month and noted seasonal changes from September to June. This is Wendy's entry for March:

3/12 2:00

1. My tree's trunk has gotten smaller.
2. My tree has a lot of moss on it.
3. My tree has a lot of snow near it.
4. My tree's branches are very bare.
5. My tree has a lot of twigs near it.

In response to Wendy's observation, we had a lengthy class discussion about how she knew that her tree's trunk was smaller. She explained that she could tell by looking. Other students pointed out that that could be inaccurate, and she decided to measure the trunk next time.

A comparison of Dixie's October and March entries shows how she captured the changing of the seasons.

10/4 2:00

Now the tree has more of a greyish color on it. It has three woodpecker holes. Also the tree looked a little bit bigger.

3/12 3:00

1. It changed color.
2. It does not have any leaves left.
3. My tag fell off.
4. My tree got a lot, lot bigger.
5. I have lichen on my tree.

We also kept a caterpillar diary in which we recorded observations of a monarch caterpillar that we kept in the classroom as it formed a chrysalis and hatched into a butterfly. I initiated the activity by reading aloud *The Caterpillar Diary* by David Drew, an excellent example of a diary of observations. The following day I read *A New Friend for Morganfield* by Ann Hobart, and we guessed what might happen to our chrysalis in response to Hobart's prediction of a fourteen-day incubation period. As a class, we discussed the difference between observations and predictions. Then I set up a class diary and students took turns writing an entry each day.

DAY 1

The caterpillar has made his chrysalis. It is green and it is shiny. It has little gold dots on the chrysalis. In a while it will turn into a monarch butterfly. Justin

Next year, I plan to have everyone keep his or her own diary, with daily sharing of individuals' observations to provide more opportunities to practice keeping track of observations.

We also kept a log of daily weather observations throughout the year, noting which data we could observe with our eyes and which we needed to measure. Observing, record keeping, and categorizing are all important procedures that researchers frequently use.

This variety of methods—note taking without copying, research notebooks, acrostics, interviews, and logs of observation—assisted third graders as they began to learn what researchers do. A combination of hands-on approaches and writing activities proved much more accessible to eight- and nine-year-olds than formal report writing.

AS THE year progressed, I allowed students more choice in formats for final products, now that they had some experience as researchers. As a class, they brainstormed a list of potential formats; they came up with seven, and I added other alternatives that I thought were appropriate for beginning researchers from the list generated by the children and teachers in the writing-to-learn project (see sidebar). Individual students then chose the formats that allowed them to share their knowledge of their sea creatures. These five were their favorites.

DIORAMAS

The diorama was one of the most popular formats because it involved painting. (Next year, the boxes will be painted during art class, since they took three or four class sessions to complete.) Each diorama included a written description of the sea creature portrayed. We had dioramas on penguins, the octopus, and killer whales.

PUBLISHED BOOKS

Other students wanted to write up their research as a story to publish in a book, complete with covers made

from cardboard and contact paper, with pages bound on a sewing machine. For this project, my colleague Jo Haney developed an improvised galley sheet: a mimeographed form with six or eight lines at the bottom that helped students divide the drafts of their stories into separate pages to be typed and illustrated. Figure 1 shows how Mark used a galley sheet to set up one page of his book about whales and to organize his information.

FIGURE 1
ONE OF MARK'S GALLEY SHEETS FOR HIS BOOK

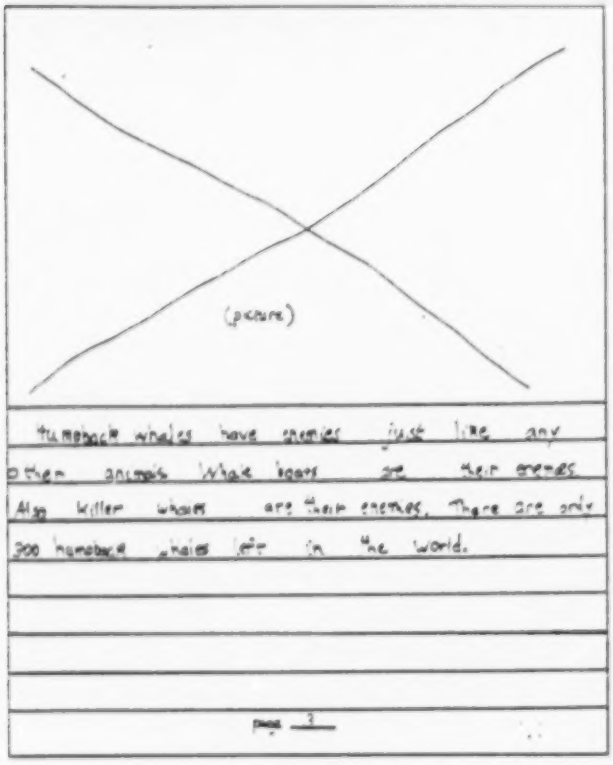
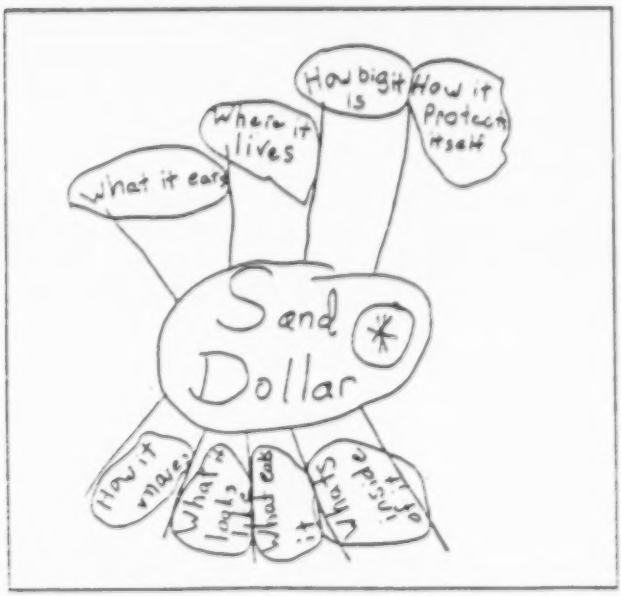

(picture)
humpback whales have enemies just like any
other animals. Their bodys are their enemys.
Also killer whales are their enemys. There are only
300 humpback whales left in the world.
page 3

FIGURE 2
DIXIE'S WEB: PREPARATION FOR HER BULLETIN BOARD



BULLETIN BOARDS

Dixie organized her research on sand dollars as a bulletin board display with a picture of a giant sand dollar in the middle and relevant facts arranged on cards around it. She planned her bulletin board in her learning log by drafting a web to help her make decisions about her topic and organize her information (Figure 2).

GAMES

Two members of my class developed games from their research. One board game was on the narwhal and took players on a trip around the Arctic, posing questions to be answered along the way. Alewives were the subject of another board game and involved a journey up the falls; players' responses to questions determined whether they climbed the alewife ladder.

TIME LINES

As part of our science curriculum, we study space. While the other two third-grade classrooms prepared a tour of the solar system, sharing facts on each planet, my students and I constructed a time line of the creation of the solar system. I read aloud *Life Story* by Virginia Burton, an account of the creation of the universe written as a play, and showed several examples of time lines, one on U.S. presidents and another on inventions. Each student selected one era that he or she would take responsibility for in creating our time line. They researched their eras, then drew pictures portraying what they had discovered about each period. Then they drafted descriptions of what was happening in their pictures. The following excerpts show the very beginning of the time line.

> Today scientists think that 4.5 billion years ago, our solar system used to be dust and gases. All of a sudden a star exploded and then our solar system was a large, spinning disk. Mike

> Two to one billion years ago the crust of our earth started to form. The earth started to wrinkle. High mountains and low valleys started to form. There were deep ocean basins. Metamorphic rock started to form. The metamorphic rock formed by the heat and pressure. There were thick clouds that blocked the sun from shining on earth. Rae

> About thirty to forty million years ago the climate started getting cooler. Mammals started developing. Plants started blooming. Volcanoes began to erupt, and mountains started to rise. Meggan

We practiced reading the time line aloud and then presented all twenty-six feet of it during an assembly for grades four to six. The idea for this particular project began with a read-aloud, was researched through independent reading, and developed first as an illustration, and then as a written description.

After our presentation, I discovered a book called *Earth Story* by Eric Maddern, a poem about the creation of the earth. As I began reading it to my class, Mike commented, "That's just like what Herbie did in his picture," and as I continued reading, eyes lit up around the room. The children recognized "their" information and were

BEST COPY AVAILABLE

proud of what they had accomplished. When I closed the book, Heather wondered why the poem had not included the creation of life, as our own time line had. Joey's hand shot up. "Maybe he can write another poem to show that!"

My students shared their projects with pride and confidence. During the preparation of final projects, I highlighted many techniques in conferences and mini-lessons: galleys, the use of pen or marker (as opposed to pencil) on bulletin boards, diorama construction, and techniques for making books. When the projects were completed, I asked the students to evaluate themselves and to write advice they might offer to next year's class. Because they had taken responsibility for their own learning, they could reflect on and assess what was helpful to them as beginning researchers and what was not.

Throughout the third-grade year, I plan activities that help students explore, investigate, experiment, and record—not in isolation but in cooperation with a teacher and classmates who offer encouragement and questions. I provide as many opportunities and as much time as I can for students to try the procedures I demonstrate.

When Nancy Nash talked with me that afternoon about her students' research, she was affirming our hard work of the year before. As a teacher, I can do something that no commercial program can ever accomplish: allow students' individual interests to become a catalyst for their involvement in their own learning. I can teach research principles and procedures in the context of classroom discussions, activities, and resources. I can expose third graders to rich print resources as well as films, tapes, and local experts and provide opportunities for them to draw on these resources in investigating self-selected topics. I can make sure that their first research project does not become an overwhelming task but rather one approached with confidence and genuine curiosity as they investigate, organize, and share new information in a variety of formats. And I can act as their guide: not expecting students to possess the skills of researchers as if by magic but planning activities, demonstrating techniques, and gathering resources that will help them to take their first steps as beginning researchers.

References

Burton, Virginia Lee. 1963. *Life Story.* Boston: Houghton Mifflin.

Dethier, Vincent. 1981. *Newberry: Life and Times of a Maine Clam.* Camden, ME: Down East Books.

Drew, David. 1987. *Caterpillar Diary.* Crystal Lake, IL: Rigby Education.

Hobart, Ann. 1985. *A New Friend for Morganfield.* Portland, ME: Gannett Books.

Johnston, Patricia. 1985. "Writing to Learn Science." In *Roots in the Sawdust,* ed. Anne Gere. Urbana, IL: National Council of Teachers of English.

Maddern, Eric. 1988. *Earth Story.* New York: Barron's.

Milton, Joyce. 1987. *Marching to Freedom: The Story of Martin Luther King, Jr.* New York: Dell.

Myers, Arthur. 1981. *Sea Creatures Do Amazing Things.* New York: Random House.

Zoobooks. Wildlife Education, Ltd., Box 28870, San Diego, California 92128-9848.

Prenatal Exposure

(Continued from page 33)

result in low birthweight due to either preterm birth or intrauterine growth retardation (Kaye et al., 1989; MacGregor et al., 1987; Petitti and Coleman, 1990). While some symptoms may be immediately visible, others may not be apparent until later childhood (Weston et al., 1989; Gray and Yaffe, 1986; Frank et al., 1988).

In infancy, damaged babies can experience problems in such taken-for-granted functions as sleeping and waking, resulting in exhaustion and poor development. In childhood, problems are found in vision, motor control, and in social interaction (Weston et al., 1989). Such problems may be caused not only by fetal drug exposure but also by insufficient prenatal care for the mother or by an unstimulating or difficult home environment for the infant (Lifschitz et al., 1985).

WHAT CAN be done to ameliorate the condition of children born with such damage? Quite a bit, based on the success of supportive prenatal care and the results of model projects that have provided intensive assistance to both baby and mother from the time of birth. These projects have successfully raised the I.Q. of low- and very-low birthweight babies an average of ten points or more—an increase that may lift a child with below-average intelligence into a higher I.Q. category (i.e., from retarded to low average or from low average to average). Generally known as either educational day care or infant day care, these programs provide a developmentally stimulating environment to high-risk babies and/or intensive parent support to prepare the parent to help her child.

In one such program based at the University of California/Los Angeles, weekly meetings were held among staff, parents, and infants over a period of four years. By the project's end, the low-birthweight babies had caught up in mental function to the control group of normal birthweight children (Rauh et al., 1988). The Infant Health and Development Project, which was conducted in eight cities and provided low-birthweight babies with pediatric follow-up and an educational curriculum with family support, on average increased their I.Q. scores by thirteen points and the scores of very-low birthweight children by more than six points. Another project targeted poor single teenage mothers whose infants were at high risk for intellectual impairment (Martin, Ramey and Ramey, 1990). One group of children was enrolled in educational day care from six and one-half weeks of age to four and one-half years for five days a week, fifty weeks a year. By four and one-half years, the children's I.Q. scores were in the normal range and ten points higher than a control group. In addition, by the time their children were four and one-half, mothers in the experimental group were more likely to have graduated from high school and be self-supporting than were mothers in the control group.

These studies indicate that some disadvantages of poverty and low birthweight can be mitigated and intellectual impairment avoided. The key is attention to the cognitive development of young children, in conjunction with social support of their families.

Genres for Report Writing

By Nancie Atwell

THIS APPENDIX presents most of the options for reporting knowledge across the disciplines generated by the children and teachers in the writing-to-learn project. Teachers who plan to invite multi-genre reports might wish to review these modes and select those most appropriate for their students and subject areas. It is not a list to hand to children but a starting point for the teacher who is considering options to present to students and is willing to show children, in conferences and mini-lessons, how the genres work.

1. *Individual bound books* for the classroom library. Giacobbe's bookbinding technique, described in Graves, is one that children can manage independently from around third grade.

2. *Picture books* that introduce younger children to a topic and are based on students' knowledge of good, content-area literature for children (e.g., illustrated books about electricity, black bears, local architecture, the human skeleton).

3. *Textbooks* for which each student in the class writes a chapter (e.g., the results of statistical surveys conducted by students in a math class, an anthology about life in ancient Greece, an examination of the effects of World War II on the local community).

4. *Correspondence* between two real or imagined historical personages (e.g., a woman from ancient Sparta and one from Athens, Thomas Paine and a twentieth-century fifth grader, Harriet Tubman and a young slave).

5. *Journals or diaries* of real or imagined historical personages (e.g., the diary of a serf, the journal of a young survivor of the flu epidemic of 1918).

6. *Oral histories and interviews*, transcribed and supplemented by background information, photographs, drawings, poetry, etc. Linda Rief's eighth-grade study of aging is a lovely example, as are oral histories published in the *Foxfire* collections edited by Eliot Wigginton.

7. *Scripts:* radio and television plays to be tape recorded or videotaped; speeches, plays, and skits to be performed; interviews; and film strips.

8. *Historical fiction:* short stories about historical personages or about imagined people taking part in important historical events (e.g., a day in a child's life during the plague or on a wagon train, a fictional account of Anne Hutchinson's trial).

9. *Autobiographical sketches* of real or imagined historical personages or living things (e.g., a first-person account of the boyhood of Alexander the Great, a deciduous tree describes a year in its life).

10. *Poetry:* collections of poems about a topic—free verse, rhymed, counted syllable and/or acrostic formats—in which information about a topic is embedded.

11. *Animal stories:* a favorite genre of third through fifth graders; the stories must strike a balance between presenting the animal as a character and giving an accurate account of its existence without anthropomorphizing it (see Wilde 1988).

12. *How-to books* in which students pass on specialized knowledge related to a unit of study (e.g., blacksmithing, trapping, tapestry weaving, stargazing, reducing fractions).

13. *Field guides* that describe characteristics of a particular species or community.

14. *Class or individual newspapers* in which each article, column, advertisement, editorial, interview, want ad, and cartoon is related to a time and place in history (e.g., a Boston newspaper of 1776, a Gettysburg paper from 1863).

15. *Games and puzzles* that demonstrate and require a knowledge of a time, place, or unit of study (e.g., a trivia game about Portland, a crossword puzzle with the solar system as its theme).

16. *Annotated catalogs* of artifacts (e.g., the dress of men and women of ancient Greece; cooking implements found in the kitchen at Sturbridge Village).

17. *Annotated family trees* of real or imagined historical personages (e.g., Greek gods and goddesses, a passenger on the *Mayflower*).

18. *Bulletin boards* of drawings or photos with accompanying text (e.g., plants that grow in the desert, Portland then and now).

19. *Posters, murals, time lines, and mobiles* that include text (e.g., a dinosaur mobile, a mural depicting the destruction of Pompeii, a poster showing a plant's life cycle).

20. *Coloring books* with accompanying text, to be photocopied for classmates and/or younger children (e.g., scenes from New England states, the Underground Railroad, the life of a hermit crab).

21. *Calendars*, each page annotated with a drawing and text related to the topic (e.g., a Medieval knight's calendar, a calendar for stargazers, a puffin calendar).

22. *Alphabet books* in which each letter supplies relevant information about the topic (e.g., a Beverly Cleary ABC, an astronaut's ABC, a geologist's ABC).

23. *Shadow boxes and dioramas* with accompanying text (e.g., the habitat of the eastern panther, Anne Frank's secret annex, the parts of a stem).

References

Graves, Donald H. 1983. *Writing: Teachers & Children at Work*. Portsmouth, NH: Heinemann.

Rief, Linda. 1985. "Why Can't We Live Like the Monarch Butterfly?" In *Breaking Ground: Teachers Relate Reading and Writing in the Elementary School*, ed. Jane Hansen, Thomas Newkirk, and Donald Graves. Portsmouth, NH: Heinemann.

Wigginton, Eliot, ed. 1972-1986. *The Foxfire Books 1-9*. Garden City, NY: Anchor Press, Doubleday.

Wilde, Jack. 1988. "The Written Report: Old Wine in New Bottles." In *Understanding Writing: Ways of Observing, Learning, and Teaching*, 2nd ed., ed. Thomas Newkirk and Nancie Atwell. Portsmouth, NH: Heinemann.

Adapted with permission from Coming to Know. *Nancie Atwell directs the Center for Teaching and Learning, a model elementary school in Edgecombe, Maine. Her book* In the Middle: Writing, Reading, and Learning with Adolescents *won the highest research award given by the National Council of Teachers of English.*

BEST COPY AVAILABLE

SECTION 4

ASSESSMENT OF SCHOOL BOARDS AND STAFF

Reprinted with permission from MASC Journal, Fall 1990.
(Originally appeared in the Journal of the Ohio School Boards Association.)

A School Board's Need for Self-Assessment

by Donald Overly and Darrell K. Root

School boards need to examine what they do and how they do it when performing board responsibilities to achieve and maintain a positive image.

It is exciting to work with boards that function in harmony and frustrating to work with those whose members choose every opportunity to disagree with each other or the school administration. When a school board operates in open conflict, the community develops a negative impression of it; efforts at any program improvement cease, and students' education is impaired.

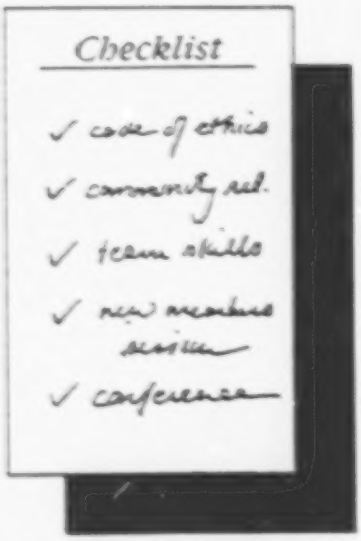
Although the majority of boards are positive and well-intended, conflict prevails on many educational scenes, especially among members of the board of education and between board members and the superintendent. Unfortunately, these tensions continue despite efforts by state school board associations to promote harmony and unity.

Open conflict emphasizes the need for self-assessment by board members to help them to understand the need to present a united front and positive image to staff and community members.

This is not to suggest, however, that all board members must vote yes on every issue or that all votes be unanimous for the board to appear unified. Philosophic differences do and will exist.

Members who know the issues can sell their views without presenting a negative image of the board and the school district. In other words, board members must disagree agreeably.

Although conflict is more likely to occur when finances are low and debts are high or when changes must be made, there are some ongoing self-assessment practices that can help board members work as a team to do what is best for students. The following self-assessment list will not guarantee harmony, but it will help minimize conflict, and, in turn, build a positive image of the school board in the community.



Prospective board members' orientation

Boards often leave orientation sessions to the district superintendent. There is nothing wrong with the superintendent's active involvement in orienting prospective board members, but present board members should take a leadership role in this process.

Board interaction with prospective candidates can pave the way for positive campaigns. During the campaign, emphasis can be placed on ethical values and student needs rather than petty personal interests.

Reorientation of new board members

A reorientation session immediately following the election of new members is vital. It should be focused on policy development, decision-making processes, and establishing a relationship with the administration.

Common problems facing many boards, such as what to do when a community member contacts individual board members also should be discussed. This reorientation process not only gives direction to new members, it reminds veteran members of the "proper" way to conduct board matters.

Code of ethics

Every board should have an adopted code of ethics. Unfortunately, most boards spend little time examining principles, procedures, and standards.

A thorough examination of them should be held annually, with particular attention given to the board's role versus the administrator's role; statements promoting communication and understanding between the board and the community; encouraging and accepting opinions from all segments of the community; establishing procedural steps for resolving complaints and criticism of school affairs; and recognizing compromise as a way to resolve conflict.

Develop effective skills

Devote several work sessions to developing team skills. Concentrate on effective problem-solving skills to help the board identify, clarify, and resolve personal and group problems. A skilled counselor/consultant could show the school board how to eliminate self-defeating behavior by using conflict-management and resolution techniques.

Adopt win-win characteristics

People need positive experiences to develop self-understanding and promote a framework in which no one loses. A session could have board members share personal thoughts, feelings, likes and attitudes. This process would enable members to identify similarities and differences that exist among people and understand each other's point of view. The final result would be to make it easier to compromise. Be aware of the change process. Change is necessary for school boards to remain effective. Board members should strive for purposeful change that demands commitment. This requires a formal process.

Organizations and conferences

Board members are often reluctant to attend state and national conferences because of cost or distance. Also, they fear the community will criticize them for their "junketing."

State and national conference formats were designed to help resolve destructive board conflict. As elected representatives, board members have an obligation to attend such meetings. To avoid criticism of impropriety, the board should pass a resolution at the organizational meeting stating that board members will attend professional meetings and participate in them whenever possible.

Expectations for the future

There will always be conflict among individuals and groups within an organization. Accepting this fact, school boards that practice self-assessment and internalize an understanding of management and team effort can be successful.

The entire school community will gain from such a joint partnership, and a more positive relationship will be achieved. Effective school boards unite communities, and this benefits the children whom board members were elected to serve. □

Dr. Overly and Dr. Root are professors in the Department of Educational Administration at the University of Dayton, Ohio. This article originally appeared in the Journal of the Ohio School Boards Association.

Reprinted with permission from NASSP Bulletin, September 1990.

Self-Assessment in the Evaluation Process

By Michael Koehler

Here's a process that enables supervisors to incorporate teacher self-assessments into their observation reports. The majority of teachers are capable of such activity, and having done so are more willing to engage in follow-up activities, according to this writer.

Supervisors seem to have accepted the notion that formative observation should constitute the bulk of their work with teachers. Finding the time to complete the task remains an issue, but encouraging the professional growth of teachers by providing frequent "mirrors" of their performance is widely accepted in most of the literature. It has not, however, found its way into most school systems with the same enthusiasm.

This fact is not too surprising: No change in education is possible without the forces *for* change being greater than the forces *opposed* to it. Systems are resistant to change. That formative observation is the best way to improve teacher performance may not be as widely accepted as much of the literature would have us believe.

Michael Koehler, a former school administrator, is with Ideation, Inc., in Mundelein, Ill.

87

What must supervision do to strengthen the forces for such change in education? One immediate answer is to make the process as desirable for supervisors as it is now supposed to be for teachers.

It can be as threatening to evaluate as it is to be evaluated. Effective evaluation requires expert power, not positional power. Without a thorough knowledge of the teaching/learning process, persons occupying hierarchical positions have only the authority to observe teachers, not the competence such a responsibility requires.

They often become as uncomfortable as the teachers they observe. The least skilled find sanction in the role they occupy, not the knowledge they seek. The most skilled realize that evaluating a teacher, especially if he or she has been teaching conscientiously for several years, generally involves one of two basic assumptions.

The first is the most obvious: Schools have assumed for years that administrators have sufficient knowledge of the teaching/ learning process to make value judgments of classroom performance and to influence the professional growth as well as the employment status of teachers. Given the experience and the technical competence required of most administrators and supervisors, the assumption seems generally reliable. That they understand the complexities of teaching is consistent with the assumption. That they understand them better than all teachers, or even most teachers, is where the assumption begins to break down.

Some administrators and supervisors are the most knowledgeable people in the building regarding teaching. Most are not. If we accept that fact, then what fundamental purpose do they serve? In essence, what function can they perform to assist with the professional development of their teaching colleagues?

That's where the second assumption comes in: Many supervisors understand the teaching/learning process but reject the "I-know-more-than-you" notion. They accept the assumption that their effectiveness is enhanced when they avoid making value judgments and when they control the process by which such value judgments are made.

If an administrator restricts his or her decisions to the process by which decisions are made and makes few terminal decisions, his or her behavior will be more acceptable to subordinates and the decisions will be more effective.

Supervision involves a practical application of these ideas. Value judgments regarding teacher performance are decisions. They are among the most significant decisions made in any school. The key is to make few of them, but to control the process by which they are made. The easiest way to do this is to remove the decisional burden of responsibility from the supervisor and leave it with the teacher, where it belongs. Both the teacher and the supervisor will be happier.

Every time we evaluate a teacher, we deprive him or her of the opportunity to self evaluate. Teacher self evaluation results in ownership of performance. It's safe to assume that the

son who makes the value judgments owns them, at least initially.

The first step is to get the teacher to make them. Woodrow Wilson said that evaluation should provide light, not heat. Supervisors should encourage the illumination that comes from teacher self-assessment and avoid applying the heat that comes from traditional evaluation.

A Simple Process

The process is relatively simple. Every worthwhile classroom observation involves the collection of data, the kind that documents one or more areas of teacher performance. Teachers must be the first to analyze these data and, ultimately, to make preliminary decisions regarding what the data say about the quality of teaching.

The process, therefore, consists of four steps. The first requires the supervisor to interpret and analyze the data regarding its implications for teaching, but to avoid sharing any conclusions with the teacher. Nothing should interfere with the teacher's self-assessment.

The second step involves giving the data to the teacher. It should be done at least 24 hours before the post-observation conference. The teacher is asked to review the materials and to be prepared to discuss them when he or she meets with the supervisor.

The third step involves the meeting, a time when the supervisor seeks to elicit analyses from the teacher, discuss the teacher's reactions, and help identify implications for future teaching. This "help mate" ste[illegible] much different from those described so exhaustively in the literature.

Step four involves the supervisor's written record of the teacher's self-evaluation. If written thoughtfully by the supervisor, it should contain no value judgments, but should reflect a control of the process by which such judgments are made by the teacher. Such a process requires of the supervisor a knowledge of teaching as well as the ability to elicit appropriate responses from the teacher.

Some Examples

Following is an example of one part of the written record developed by a supervisor in a suburban Chicago high school. She submitted it for critique before giving it to the teacher:

> I enjoy the way you engaged your class in the comparison of "Dulce et Decorum Est" with the final chapter of Crane's Red Badge of Courage. The students, particularly Tom in the front row, entered into the conversation with real enthusiasm, especially when you probed his comment that there is nothing glorious about fighting and dying for one's country. I noticed two students in the back of the class—Jim and Fred—who never paid much attention to any of the discussion. It might be wise to call on them periodically or to use Silver's "Think Book" concept to get them engaged. . . .

Our critique resulted in the following modified report, which she subsequently gave to the teacher, after meeting with her to discuss the data and to encourage the teacher's self-analysis:

> You were mighty proud of that lesson! After a quick look at my script tape, you

BEST COPY AVAILABLE

decided that your decision to compare and contrast "Dulce et Decorum Est" with the final chapter of Crane's Red Badge of Courage was a good strategy. You were especially pleased with—as you put it—"the higher order thought process it provoked in the class, especially in Tom, who got right to the heart of the lesson with his comments."

I shared your position that comparison and contrast invariably leads students toward analysis and synthesis level thinking and that probing their initial observations usually engages most of the class. You felt that Cynthia's comment that the poem provoked a stronger emotional response than the book is a good example.

You were somewhat disturbed about Jim and Fred and their apparent reluctance to "dive in." You observed that they do this quite often. After some discussion, you decided to formulate a few questions— "just for them"—the kind that are likely to get a reaction from both of them. I'd like to watch the strategy at some time in the future; I want to try it in my class with a couple people who always seem to have their minds somewhere else!. . . .

After giving the observation report to the teacher, the supervisor indicated that she felt better about the whole process because she didn't have to make any value judgments. The teacher made all of them, and the supervisor merely "recorded" them. She was pleased that her skills involved helping the teacher analyze the data, not fending off attacks from a frustrated teacher who disagreed with her judgment of the class. She enjoyed the task of being a "help-mate" throughout the process.

A second example makes the point even clearer. Another supervisor, this one from inner-city Chicago, mentioned that she was working with a particularly troublesome teacher. She brought in her next observation report before giving it to the teacher.

Following is a segment of what she had written:

> . . . I've attached a copy of Madeline Hunter's observations about teaching toward an objective. Take a look at it before we meet again. I think you'll find it useful. At times this morning I found you rambling and occasionally losing sight of your objective. The instructional input regarding Wordsworth's background and how it influenced his writing was generally interesting, but at times I was wondering where you were going with it. It might be wise next time to sustain a focus on the purposes of the lesson and to relate all the instruction to each of the purposes. How, for example, does "Tintern Abbey" provide examples of Wordsworth's love of nature as developed during his youth?
>
> If you'd like to discuss any of this, stop by. I'm always available to you. . . .

Again, the supervisor was encouraged to share the data with the teacher at least one day before the post-observation conference and then to help lead an analysis of it. In addition, several questions were suggested for the conference: "Did the lesson go as you had planned?" "What do the data tell you about the instructional input as it related to your objectives?" "Did you observe anything in your students' behavior that supplements what I recorded?" "What might I have done differently?" "How can I help during future lessons?"

Certainly, none of these questions is far different from those asked by any good supervisor, but they did help to provoke the kind of dialog needed to develop a more evocative report, which included the following:

> . . . when we discussed the purposes of this lesson, I was intrigued by your comment that you have always had a tendency to ramble. As we both thought about your comment, you decided that it might be wise to specify your objectives for each lesson and then to put five or six critical questions on note cards for periodic reference during class. I asked if I could help somehow, and you asked that I visit early next week to tally the number of times you ask the questions. Madeline Hunter has some thoughts on the subject, too.
>
> I took a moment to Xerox a few of her comments. They're attached. Maybe they'll provide another dimension for your thinking. Use the materials as you see fit. If I can help in any other way, stop by. I'm always available to gather information for you. At any rate, I'll be in sometime early next week to help out. Let me know beforehand exactly what you want. . .

Teachers are more likely to accept these kinds of comments. As important, because the supervisor makes no value judgments/ decisions, his or her involvement in the process is more acceptable to the teacher. The process, therefore, is less threatening to both of them and more supportive of their respective positions in the school, the teacher as, professional and the supervisor as a knowledgeable colleague. Most important, the teacher is more likely to accept the substance of the report, to work harder to improve areas of weakness, and to seek the supervisor's help when needed.

Conclusion

Obviously, not every teacher is capable of the kind of self evaluation that leads to ownership of performance. Those who are either inexperienced or incapable require direct assistance from supervisors. Such assistance must involve specific value judgments. Such judgments should be followed by directed experiences for the teacher that remediate areas of weakness. Knowledgeable writers in the field assert that teachers requiring such direct assistance constitute only 5 to 10 percent of the typical school staff. The majority of teachers, therefore, will benefit from a process that engages them in self-evaluative activities most of the time.

Supervisors will realize similar benefits. When they incorporate teacher self-assessments into observation reports, their teachers are more willing to engage in follow-up professional growth activities and to perceive supervisors as help-mates in the process. And when supervisors are perceived as help-mates, they become immediately more effective, and their levels of self-satisfaction necessarily increase. The process is beneficial to everyone, particularly to students, who realize the greatest benefits from the professional growth experiences of their teachers.

Reprinted with permission from The School Administrator, December 1989.

How Best To Evaluate Principals?

Spotlight on Site-Level Operations Sparks Keen Interest in Principal Evaluations

BY MARK E. ANDERSON AND LINDA LUMSDEN

School principals across the country are under more pressure than ever before to show measurable performance results within their buildings.

As awareness has grown over the significant role principals play in affecting performance and attitudes of students and faculty, so too has the push for formal evaluation procedures.

While classroom teachers have been scrutinized for years, principal performance has not always been viewed as a critical link in the achievement of excellence in schools.

"Never before in our nation's striving for school changes has the focus on the principal as the primary change agent been stronger," says Richard Schmuck, writing in the spring 1988 issue of the Oregon School Study Council Report.

This focus on principal competence is contributing to keen interest in the methods and procedures school districts use to evaluate principals. Current studies suggest these evaluation practices often are inadequate and ill-suited to improving the performance and competence of building administrators.

In some instances, the evaluation of principals by central office administrators is downright abysmal, according to researchers Joseph Murphy, Philip Hallinger, and Kent Peterson.

Widespread Benefits

However, before designing and implementing a comprehensive evaluation program for principals, districts must determine the purposes of the process. Research suggests the reasons for evaluating principals can be classified under one of two general types: formative evaluation or summative evaluation.

Formative evaluation is intended to help principals improve their performance. This process requires ongoing communication between superiors and principals, with the focus on improving not only principal performance, but also the overall educational program.

Summative evaluation serves as a judgment of performance, used to guide decisions regarding promotion, demotion, incentive pay rewards, and other personnel actions.

Districts that weave both purposes—accountability and professional improvement—into a comprehensive evaluation program reap important benefits.

From summative evaluations, school districts gain public support by dismissing or reassigning incompetent principals following a rigorous evaluation. From formative evaluations, principals can improve their performance and enhance the school building's performance through meaningful communication with and effective supervision by supervisors.

Three Phases

Dale Bolton, author of *Evaluating Administrative Personnel in School Systems*, says planning and using a systematic evaluation scheme that meets the dual functions of accountability and professional improvement involve three basic phases:

- *Planning for evaluation.* This includes analyzing a specific situation, outlining the purposes of evaluation, setting goals and specific objectives, and deciding on the means for measuring the processes used and the eventual outcomes.
- *Collecting information.* This involves monitoring and measuring the activities planned and the outcomes that result from the activities.
- *Using information.* This includes communicating information that has been analyzed and interpreted and determining the next steps.

Photo by Susie Fitzhugh © 1989

> ***"It's important not only to provide principals with feedback, but to provide it in a timely fashion."***

The Planning Phase

The first step in planning an effective evaluation system involves specifying the district's values, philosophy toward evaluation, purposes of the process, and expectations of principals. School districts err seriously by failing to involve a broad base of school personnel in designing the evaluation system.

Determining and communicating what is expected of principals is the most critical aspect of planning. Yet school districts often fail to give principals clear expectations.

In their 1988 study of evaluation practices, William Harrison and Kent Peterson found many principals who were uncertain about superintendent expectations for their performance.

Daniel Duke and Richard Stiggens, in a report three years earlier, found similar confusion among principals. They concluded superiors must articulate their expectations for principal performance and "expend greater effort in explaining the evaluation system to those subject to it."

Another important planning component is the development of performance goals for the school system, school building, and individuals.

Jerry Valentine, in a speech to the American Association of School Administrators in 1986, argued that principals often set goals that were too simplistic. In his research on setting school goals for the purpose of assessing principals, Valentine found many instances where goals set by principals were merely "statements about personal administrative skills rather than desired outcomes."

He added, "The goals frequently reflected activities already completed or nearly completed and did not make a significant difference in the quality of life or program of the school."

To assist principals, Valentine recommends building administrators be given training to help them set meaningful school goals. He also suggests school faculties and principals limit the number of goals each year, recommending "two or three unique to a particular school and two or three correlated with district goals."

Collecting Information

The data-collection phase, along with communicating results, is at the crux of the formative approach to evaluation, according to Valentine. Unfortunately, many principals report they are not informed of what performance-related information is gathered or how it is obtained.

Harrison and Peterson, in their research, found only half of the principals they surveyed "claimed to know how the superintendent accumulated information upon which they based their evaluations."

Therefore, school officials must not only develop and adopt sophisticated collection strategies, but also must discuss with principals the who, what, when, where, and how of information collection.

One way superintendents can collect data is through on-site visits, in which they shadow the principal for extended periods, recording specific statements and actions.

They also can gain insight into performance by observing principals at faculty meetings, parent conferences, teacher observation and evaluation conferences, and public meetings.

Although the superintendent is perceived as the appropriate person to observe the principal, many superintendents or supervisors fail to observe principals at work.

Duke and Suggens found most evidence for principal evaluation comes only from superiors' perceptions of how principals perform, not from directly observing them.

One reason for this, they said, is that superintendents do not have time to conduct thorough evaluations. A way to increase the amount of useful information is to "involve other school personnel in collecting information on principal performance," the researchers said.

Peer observation and feedback can be useful, although the peer review process has not yet achieved wide acceptance. Information provided by teachers, parents, and students also can be used, although this form of rating principals is not common among school districts.

Another available strategy is to assess principals' skills and performance through assessment center simulations. In a typical assessment center, participants are observed completing simulation exercises by a team of trained assessors who record participants' behaviors on 12 skill dimensions.

The assessors discuss each participant and reach consensus on his or her performance. In an intensive feedback session, each participant receives a comprehensive report describing his or her strengths and weaknesses, plus suggestions for improvement.

Artifacts also can play a role in assessing principal performance. Artifacts include student test scores; written evidence of progress toward a principal's personal, school, and district goals; principal awards and recognitions; newspaper clippings,

continued on page 22

continued from page 18
parent newsletters, and letters to staff and students; and logs of how principals spend their time.

Using Information

Once information has been gathered, the evaluator must analyze, interpret, and make decisions based on it. During this phase, communication is

uators spend the first 30 minutes of conferences discussing all the 'good' skills and save 'bad' skills for the last few minutes."

Those being evaluated often tend to "tune out" the positive feedback they receive because they are anxious about or are bracing themselves for less favorable comments to follow.

The final summative reports by

Photo by Peggy Sirott Hampton

"Feedback on principal performance also should be solicited from peers and teachers."

vital.

It's important not only to provide principals with feedback but to provide it in a timely fashion. As Bolton contends, "The accumulation of information over a long period of time is not as beneficial as providing it soon after events occur."

From his research on principal evaluation practices, Valentine has found too many instances when "evaluators should culminate ongoing communication throughout the year, and not become a sudden judgment that catches principals by surprise. Principals and evaluators should identify areas for improvement and develop growth plans early in the year and then make final judgment regarding performance on each criterion that was established.

In addition, the summative evaluation should document progress made by principals toward school- and district-wide goals.

Difficult Decisions

The final aspect of an evaluation involves making a decision about employment status and pay. Valentine argues that contractual decisions always should be based on data collected during the year and summarized in evaluation documents.

"To determine employment on any basis other than the findings would be unfair to the principal," he says.

Merit pay or incentive pay also may depend upon performance evaluations. According to Duke and Stiggins, many principals consider the greatest shortcoming of their current evaluation system to be the absence of rewards for outstanding performance.

But, as the Educational Research Service cautions, "developing an incentive pay system is a complicated process, and one that, even more than a standard evaluation system, requires careful planning and consideration."

Because of the intricacies of designing and implementing a sound merit pay system, some districts may elect to provide alternative rewards for principals, such as incentives for professional development and opportunities for improvement.

In the Tigard, Ore., School District, for example, principals may choose among four types of professional growth opportunities if their performance is rated as exemplary.

The rewards are: 1) attending a national convention, 2) pursuing research in an area of professional interest, 3) purchasing membership in professional organizations, or 4) proposing an individual plan that is approved by the superintendent.

Ongoing Process

Research indicates that a comprehensive and systematic approach to evaluating principals is crucial to improving principals' performance. A successful evaluation system requires that the purposes of the evaluation process be discussed, identified, and agreed upon by all who are involved in it.

Once the purposes of evaluation are in place, districts should develop a written set of performance standards and expectations by which principal

BellSouth Takes Stock of Principals

The Excellent Principal Inventory is a unique evaluation instrument, developed by the BellSouth Corporation, to enable principals to obtain confidential feedback on their performance from superiors, peers, and teachers.

Drawn from research on effective schools and literature on leadership and management, the following five commitments reflect the values and behaviors developed for the Excellent Principal Inventory:

Commitment to:	Reflected in these behaviors:
Student Success	Demonstrating Respect for Students Pursuing All-Around Excellence
Teaching and Learning	Promoting Teaching and Learning Supporting Continuous Learning as a Lifetime Goal
The School Staff	Demonstrating Respect for the School Staff Helping Individuals Improve Building a Collegial Staff
Innovation	Supporting Creativity Supporting Upward Communication
Leadership	Demonstrating Integrity Presenting Ideas Taking Responsibility Relating to External Constituencies

Eighty-nine specific behaviors within these five areas of commitment are listed on the inventory to rate principal performance.

Gerald Bogen, an associate dean in the College of Education at University of Oregon, who refined the instrument, said principals distribute the questionnaire to teachers in their schools, peers, and superiors whom "they believe are knowledgeable about their performance, whose opinions they value, and individuals who they believe will provide honest feedback to help in their professional development."

Principals also complete the inventory to assess their own performance.

Ensuring respondent anonymity is critical to obtaining honest feedback. Respondents are coded only by category: teacher, other (peer, supervisor), or self. Only a consultant and the principal see the results.

The inventory is divided into two major parts. Section one rates a principal's attitudes and actions on the 89 specific performance indicators.

In the second section, respondents can elaborate on aspects of the principal's behaviors addressed in the first section or offer feedback on issues not mentioned before.

Respondents send the questionnaire to a scoring service that tabulates the responses and develops an individual performance profile. The scoring service provides the principal with a profile report that a trained consultant interprets to the principal during a 2 1/2-day follow-up training session.

During the feedback conference, the principal and consultant develop strategies to strengthen weak areas.

One drawback to the Excellent Principal Inventory is its expense. While BellSouth has a commitment to provide principals in nine southern states with performance feedback and training, the support of other corporations is needed to help defer costs for districts in other parts of the country.

— Mark E. Anderson and Linda Lumsden

performance will be judged.

Principals should be encouraged to set meaningful individual and schoolwide goals and to engage in self-appraisal. In addition, superintendents should observe principals through scheduled and unscheduled visits and follow up these periodic visits with post-observation conferences.

Feedback on principal performance also should be solicited from peers and teachers.

Finally, districts should adopt a cyclical approach to evaluation, making it an ongoing process guided by the three phases—planning, collecting, and using information.

Only when evaluation becomes a mutual effort between principals and those who supervise them will districts be able to ensure the improvement of our nation's schools.

Mark E. Anderson is director of employment services in the Mukilteo School District in Everett, Wash. Linda Lumsden is associate editor of the Oregon School Study Council and the ERIC Clearinghouse on Educational Management at the University of Oregon. This article is adapted from Evaluating Principals, *written by Anderson and published by the study council in April 1989.*

Reprinted with permission from NASSP Bulletin, September 1990.

A Reflective Approach

Educative Possibilities in Teacher Evaluation: Two Alternatives

By Andrew Gitlin and John Smyth

Most teacher evaluation schemes follow a similar, research-based logic. These writers argue that this approach alienates teachers and limits the possibility of change.

Purporting to modify teacher behavior, as most teacher evaluation schemes do, seems a matter of common sense. On closer inspection, however, it becomes clear that such attempts at changing teacher behavior without corresponding changes in the way teachers think about schooling, is unlikely to have an impact in any sensible way on teachers or their teaching.

This is so for several reasons. First, because school realities are constituted by the multiple interpretations, beliefs, and intentions of educational practitioners, the dominant approach of teacher evaluation schemes fail to influence the interpretive

Andrew Gitlin is in the department of education studies, University of Utah, Salt Lake City, and John Smyth is professor, Education Studies Centre, Deakin University, Victoria, Australia.

frameworks (or ideologies) through which teachers view their classrooms. What teachers see as educational problems, and what they take for granted, remain untouched when all that occurs is that they are required to act under threat of bureaucratic sanctions.

Second, when the ideologies teachers hold remain unchallenged, even supposedly helpful suggestions about teacher behavior are likely to have little effect. If, for example, a teacher assumes that it is his or her responsibility to always supply the "right" answers to students, then increasing "wait time" before students answer questions does nothing to alter the essentially authoritarian teacher/student relationship that lies behind this kind of pedagogy. There might be a behavioral change of sorts, but the philosophy that informs it remains the same.

Finally, emphasizing behavioral changes that amount to improvements in classroom efficiency focuses only on narrow educational means and leaves the valued intentions behind them unquestioned, unopposed, and undebated. When educational ends or purposes are not made part of the evaluation process, the political, ethical, and moral implications of schooling remain obscured. Concealing the ends toward which teaching and schooling are directed through dominant evaluation schemes serves to maintain and help reproduce social inequalities.

The limitation of focusing solely on behavior is co[illegible]nded by an overzealous relian[illegible] a narrowly co[illegible] [illegible] of scientific research. It somehow sounds more respectable when teacher evaluators are able to say that certain teaching behaviors are desirable because they have been legitimized by findings from research on teaching. Because "scientific" research is often assumed to produce "objective" accounts of good teaching, it is not altogether surprising that evaluation schemes which utilize the products of such research feel no obligation to consult with, or ask teachers' opinions about, suggested changes. Under these circumstances, practitioner-generated knowledge is devalued, ignored, and discredited.

> When educational ends or purposes are not made part of the evaluation process, the political, ethical, and moral implications of schooling remain obscured.

Such an exclusion of practitioner knowledge not only reinforces educational hierarchies that regard teachers as being non-expert, but it also raises teachers' ire and general distrust of evaluation processes in general. Evaluation thus becomes transformed into a kind of game where winning means finding ways to please the supervisor, and then returning to the same old routines when the door closes.

Alternative Approaches

In contrast to the dominant tradition described above, "horizontal evaluation" (Gitlin and Goldstein, 1987), and a "critical version of clinical supervision" (Smyth, 1988) attempt to promote a form of change in which teachers have the power to act autonomously with respect to questioning and making sense of their teaching. Viewed in this manner, practitioner knowledge not only becomes a central part of the evaluation process, but furthermore, it enables teachers to look critically at what they know about the nature of teaching and learning.

This examination of the practices and ideologies that surround teaching enables teachers to render a critical account of teaching, one that is opposed to acting out of habit. Rather than supposed experts passing judgments about teaching, these alternative more educative approaches are based on a dialogical relationship in which two participants work together to arrive at a negotiated assessment of teaching and what it means.

It is important to note that such a dialogical relation does not imply a unanimity of either perception or judgment, but rather relies on a desire to comprehend and understand what is occurring, along with a willingness to accept that an expressed view is only partial and sometimes even wrong. Finally, where the dominant tradition focuses on the means to assumed ends, educative approaches to the evaluation of teaching assess both the way in which teaching is enacted (i.e., the means), the valued social purposes to which it is directed (i.e., the end), and the interrelationship of one to the other. What this does is ensure that, in the evaluation of teaching, the political, moral, and ethical implications of that teaching can be exposed and opened to question.

Both horizontal evaluation and the critical approach to clinical supervision are part of a growing democratic tradition of educative evaluation that emphasizes both active teacher involvement in the process and a more expansive view of what can be the object of scrutiny. One of the primary commitments underpinning this tradition is that teachers' understanding of themselves is a first step in their radically altering the self-destructive patterns of interaction that characterize social relations. An educative model, as we construe it, considers how understanding and behavior are intimately linked and enable teachers to decide for themselves, on the basis of lucid, critical self-awareness, the manner in which they wish to live (Fay, 1977).

As the dialog implicit in this tradition spreads to incorporate other teachers, and those outside classrooms who have a legitimate interest in what transpires in schools and classrooms, then collective dialog and action become both more inclusive and more possible as those structures that restrict schooling to utilitarian purposes are confronted and supplanted.

Horizontal Evaluation

Horizontal evaluation is a process in which participant tea[illegible] start out by collaborating to analy[illegible] the relation

ship between their teaching intentions and their practices, in ways that point to contradictions. This is another way of saying they are searching for the gap between what they desire to do in their teaching and what they actually end up doing.

Where there is not this mismatch between intention and practice, teachers think through why it is they want to achieve particular ends they have identified, as opposed to unquestioningly accepting others. These intentions can be stated in advance, or they can emerge from discussion.

When stated in advance such intentions enable the examination of the values implicit in what the teacher hopes to achieve. A teacher, for example, might have the intention of having all students obey the rule "raise hands before speaking." Instead of simply accepting this and observing the extent to which practice reflects this, it is important that teacher and colleague discuss why this rule is important, under what conditions, and when it might be most appropriate. Once issues like this are clarified, their desirability can be examined and debated in relation to a normative framework.

Horizontal evaluation draws primarily on the work of Gadamer (1975) and Habermas (1976) as a way of being clearer about the relationship between educational means and ends. "Communication analysis" and "historical perspective" are important aspects of the discussion that occurs between teacher and observer.

- **Communication Analysis**

Communication analysis involves a structured way in which participants examine how the prejudgments they hold about teaching help to form and frame their teaching. For example, if a teacher regards chaotic student behavior as being caused by the "low" abilities of the students, the question could be posed "What do you mean by low abilities?" Taken-for-granted values that lie behind statements and practices like this can, therefore, be made explicit and reconsidered by the participants.

Reconsideration of these values can amount to a transformation in the way teachers make and remake school reality on a daily basis. If teachers realize that the labeling they use is inappropriate because it is based on an undisclosed view of the socioeconomic status of the students, then classroom structures and pedagogy can be reorganized to reflect a more egalitarian view of the way students should be treated.

- **Historical Perspective**

Adopting a historical perspective with respect to particular teaching strategies allows apparently common sense notions and actions to be seen not as natural and immutable, but rather as choices which are part of a historical tradition that serves particular interests, while denying others. For instance, if a teacher regards teaching as a process of depositing information in the heads of students, adopting a historical perspective can encourage discussion about what factors have historically encouraged this kind of pedagogy. The interests embedded within such a view can then be more fully analyzed.

- **Challenge Statements**

Where discussion between teacher and colleague about in-class activity has become stilted, stalled, or bogged down and cannot go beyond clarifying values and prejudgments, then either of the parties can initiate a challenge statement designed to get discussion moving again. For example, if a teacher holds to the view that students should obey classroom rules in all circumstances, questions can be raised about students' and teachers' rights, and about the real role and purpose of adult authority. The resulting conversation might seek to probe the legitimacy of adult authority and the implications this holds for the education of students in a democracy.

Critical Approach to Clinical Supervision

Most readers will already be familiar with the pioneering work of Robert Goldhammer (1969) and Morris Cogan (1973), which proposed a classroom-focused, collegial model of teacher evaluation that relied on interclass visitation, data gathering, and collegial conferences. Readers will probably be familiar with this process because of its much-publicized procedural aspect.

Yet, what has received much less publicity, for understandable reasons, is the way in which the democratic relations Goldhammer had in mind have actually been twisted, distorted, and tarnished out of all recognition (Hunter, 1985). What started out as a process that had the potential to enable teachers to develop the kind of educative relationships we have spoken about here has degenerated into nothing more than an autocratic form of inspection, surveillance, and quality control that is imposed by administrators upon teachers.

> Both horizontal evaluation and the critical approach to clinical supervision are part of a growing democratic tradition of educative evaluation that emphasizes both active teacher involvement in the process and a more expansive view of what can be the object of scrutiny.

In more recent times, however, there have been some concerted moves to return to the original ideas of clinical supervision and to build on them in a way that is much more enlightening, inclusive, and liberating for teachers.

Within this approach to teachers evaluating what they do, the *procedural* stages of preobservation conference, observation and data collection, post-observation conference, and post-conference analysis, therefore, become less important than what occurs *substantively* in the process of challenging and making sense of teaching. These

substantive phases, described below, owe much to the problem-setting and problem-posing work of Brazilian educator Paulo Freire (1972, 1985).

- **Problematizing Teaching**

The emphasis here shifts from one of supposed universal certainties based on research in teaching, to problematizing (or bringing into question) the production, circulation, and consumption of educational knowledge. The collegial conferral framework of clinical supervision involves the teacher and an observer using "public space" (Greene, 1985) prior to actual teaching, to deliberately plan and rehearse how to inject into teaching, questions that normally go unasked about the content as well as the methodology of what is to be taught.

Adopting a problematizing approach to teaching introduces a dimension that opens up for debate the preprocessed and predigested nature of such knowledge by questioning the way in which powerful and pervasive myths are used to explain the wider inequalities that exist in society.

Conferral between teachers in clinical supervision prior to their actual teaching provides a safeguard whereby they are able to sketch out the limits of their teaching situations prior to the enactment of the lesson. Rather than feeling that they have to go it alone in situations of curriculum exploration, it is here that teachers begin to develop the kind of solidarity and teaching alliances with which to progressively share the problem of redefining the content of the[illegible]aching. Teachers are able to see m[illegible]learly the political nature of their teaching, and the disabling consequences that accompany such bureaucratic intrusions into the classroom as mandated curricula, prescribed textbooks, "bench mark" testing, and the like.

- **Observing and Creating Text**

Creating a written descriptive account of what actually occurs during teaching provides teachers with an account they can use to evaluate the relationship between what transpires in their teaching and the wider cultural and political life that shapes and informs schooling. Without that account, limited and selective though it may be, it is not possible to recall teaching in terms that enable its richness and complexity to be adequately addressed.

Description is an important starting point in teachers owning and analyzing their teaching. Before teachers can confront themselves and their teaching in the sense of asking poignant questions about the nature and consequences of their classroom practices, they first need to have a somewhat stable image of what is occurring.

Being clear about the elements (the who, what, when, and where) of their teaching means that they are more likely to have a basis on which to consider how their consciousness was formed in the first place, and what political conditions continue to limit and constrain them from working differently in the future.

- **Confronting Biography and History**

Teachers continually struggle to survive in a world that has certain purposes, desires, and needs, but where their capacity to satisfy them is continually being shaped and reshaped by constraints beyond their control. The challenge, therefore, is to struggle to disorient and confront what is considered normal or natural. This is only possible when teachers can put themselves at some distance from their teaching and begin the process of locating themselves in it.

At a practical level, this might mean teacher and colleague asking the following kinds of questions in the pre and post-lesson conferrals:

— Why do we insist on external rewards and punishments to make kids learn?

— Why do we define the good kids as the quiet kids?

— Why do we insist on equating workbook work with reading?

— Why do we regard on-task time as synonymous with learning?

— Why do we accept that getting through the material should be the prime goal of teaching? Indeed, where did this notion come from, and how is it related to efficiency in factories?

It is questions like these that put the focus on the "why" of teaching; that enable teachers to uncover the awesome extent of the habitual in teaching. Some of these are notions that have widespread and unquestioned status in teaching. Examining them involves an element of discomfort, even threat, but reality for most of us is that unless we feel uncomfortable, shaken, or forced to look at ourselves and our circumstances, there is little likelihood of that happening.

- **Refocusing and Action**

It is not sufficient, of course, to merely have reflected in a speculative way about the oppressive and disabling nature of classrooms, learning, and teachers' work; there must be an actual commitment to changing those circumstances. To be a truly reflective process, clinical supervision must also engage the participants in looking inquiringly at the process of clinical supervision itself.

References

Cogan, M. *Clinical Supervision*. Boston, Mass.: Houghton-Mifflin, 1973.

Fay, B. "How People Change Themselves: The Relationship Between Critical Theory and Its Audience." In *Political Theory and Praxis: New Perspectives*, edited by T. Ball. Minneapolis, Minn.: University of Minnesota Press, 1977.

Freire, P. *Pedagogy of the Oppressed*. Harmondsworth: Penguin, 1972.

———. *The Politics of Education: Culture, Power and Liberation*. South Hadley, Mass.: Bergin & Garvey, 1985.

Gadamer, H. *Truth and Method*. Translated by G. Barden and J. Cummings. New York: Seaburg Press, 1975.

Gitlin, A., and Goldstein, S. "A Dialogical Approach to Understanding: Horizontal Evaluation." *Educational Theory* 1(1987): 17-27.

Goldhammer, R. *Clinical Supervision: Special Methods for the Supervision of Teachers*. New York: Holt, Rinehart & Winston, [illegible]

Reprinted with permission from The Executive Educator, April.
Copyright 1992 the National School Boards Association.
All rights reserved.

WATCHING TEACHERS TEACH

By Judith D. Clemente, Ruth B. Greenblatt, and Robert E. Maher

Instructional improvement depends on a dialogue between the principal and the teacher

PRINCIPALS SHOULD RECOGNIZE THAT IMPROVED instruction requires teacher observation and feedback. In fact, observing teachers and then advising them on ways to improve their teaching methods are the keys to becoming an instructional leader—and creating an effective school. Only principals who have a first-hand knowledge of teachers' instructional techniques can initiate the kind of dialogue that leads teachers to self-analysis and self-improvement.

We can almost hear you grinding your teeth and running for the aspirin bottle. "Are you crazy?" you say. "How can you expect us to become instructional leaders when we are so busy building coalitions, managing, disciplining, scheduling, budgeting, and simply trying to survive?"

True, the responsibilities of the principal are ever increasing. Yet we contend that, as a principal, your most important role is establishing yourself as an instructional leader. Gone are the days when principals could serve simply as building managers. To ensure that the quality of instruction is top notch in your school, you must "lead" teachers by placing a priority on observation and feedback.

How do you do that? Visibility, for starters—you have to be seen in the classroom. But as millionaire and publisher Malcolm Forbes warned, "Establishing a visibility, a presence, is more than just being there." In other words, proving your concern about instruction takes more than spending time in the classroom. You must also roll up your sleeves and work hard with teachers to help them reach their potential.

Effective principals design comprehensive observation and evaluation plans that fit their needs. For example, you might find it appropriate to observe each teacher five times a year—often enough to give you a chance to get a real sense of that teacher's instructional style. True, for a staff of 50, that means 250 observations a year, but such commitment is critical for an instructional leader. After all, you must spend enough time with individual teachers to evaluate their efforts fairly.

How should you go about these observations? We recommend you observe each teacher for at least 30 minutes, then write a brief summary of your perceptions. Give a copy of your summary to the teacher by the end of the day. Encourage the teacher to ask questions about the observation results—and to come to some agreement with you. Alas, sometimes you must agree to disagree. But in any case, try to focus the discussion on the real issue—what is best for the students' education. Then, at the end of the year, use the observation results as the basis for writing an annual evaluation.

The result of regular observations is that teachers have a good sense of how they are doing. In addition, the observations give you a starting point for peer mentor programs and staff development and instructional improvement efforts. None of this will be easy. But if you're willing to try, consider these practical recommendations:

- ***Make classroom observations a priority.*** Let everyone know the importance you place on classroom observations. Tell your secretary and assistant principal to avoid interrupting you during observations. Talk about the importance of observing teachers. Eventually, everyone will look to you for instructional leadership.
- ***Design an observation system that works for you.*** Develop an observation schedule that puts you in the classroom for at least one-third of the school day. We suggest observations be informal, unannounced visits lasting from 20 to 30 minutes. Making at least five observations per teacher each year will make it less likely that you observe a teacher only on an especially good or bad day.

Judith D. Clemente *is supervisor of special education alternative high schools and programs for the New York City Board of Education.* ***Ruth B. Greenblatt*** *is director of the Lower Hudson Principals' Center, Elmsford, N.Y.* ***Robert E. Maher*** *is principal of Cornwall High School, Cornwall-on-the-Hudson, N.Y.*

BEST COPY AVAILABLE

• *Offer teachers immediate feedback.* The goal of classroom observations is instructional improvement, so tell teachers your reactions as soon as possible. Mention what you liked about the lesson, and talk about what you didn't like. Teachers cannot tell what they're doing wrong—or right—unless you tell them.

• *Be generous with praise, specific with concerns.* As instructional leader, you must acknowledge teachers who are doing a good job. Commenting on good instructional techniques not only boosts morale, it also might alert teachers to actions they were not aware of. It's equally important to provide specific suggestions for improvement when it's needed. Simply telling teachers they aren't maintaining classroom order is useless. But giving those teachers classroom management strategies they can use to stop students from talking back can lead to positive change.

• *Monitor teachers with problems.* If you identify a teacher with a problem, you cannot simply inform the teacher and forget the matter. You must monitor the teacher's efforts to resolve the problem. You must be ready to offer additional advice, coaching, or staff development to help the teacher. Then you must continue to observe the teacher in search of noticeable improvement.

As principal, you must establish your own model for instructional supervision. But two things are certain: Instructional improvement happens only when dialogue takes place between the principal and the teacher about classroom instruction, and that kind of dialogue means the most after classroom visits and observations. ☐

THE SUPERVISION QUANDARY

By Andrea R. Stein

Supervision and evaluation are not one and the same, but sometimes people lose sight of the distinction

GOOD PRINCIPALS DEVOTE A LOT OF TIME TO the supervision and evaluation of teachers. In fact, they often see these two responsibilities as one and the same—but they're not. The distinction is important, and principals must handle both successfully if they expect teachers to be effective.

First, let's distinguish between the two: Evaluation is the act of determining whether a teacher is doing a good job or not and to what degree. In evaluation, you are rating performance. Supervision, on the other hand, is the act of determining what's happening in the classroom. You are not rating a teacher; you are simply keeping informed and, perhaps, giving a little advice.

From personal experience, I've found that fulfilling these two responsibilities isn't easy. Consider the supervision of teachers: When I became principal of Como Park Elementary School (K-5; enrollment 500) in Lancaster, N.Y., I began dropping into classrooms for brief visits. I'd sit in the back of the room and listen, or I'd walk around to see what children were doing. I made an effort to leave a note for the teacher, praising an instructional technique I liked or making a suggestion that might help.

Andrea R. Stein is principal of Como Park Elementary School, Lancaster, N.Y.

Supervision or evaluation?

That's when I discovered the confusion that many people have in differentiating between supervision and evaluation. Within the week, I was visited by the teacher association representative, who stated I was evaluating teachers and, in doing so, not following guidelines established contractually.

I disagreed. I was supervising, I said. As principal, it was my job to visit classrooms to see what was happen-

ing. Leaving teachers a note didn't mean I was evaluating them. I was leaving advice, which I would be glad to discuss if teachers approached me. But, I said, I was giving suggestions on a take-it-or-leave-it basis.

Our disagreement lasted several months. I told the association I would continue to visit classrooms, and in fact, I made myself extremely visible. The problem came to a head when the teachers met without me. They argued: Some teachers said I was doing what should have been done all along, and others said I should announce my visits before dropping in on a class.

The superintendent, I might add, was well aware of what was going on in my building. He and the school board backed me 100 percent. They believed my supervisory efforts were a basic right and responsibility of a school principal. And ultimately the teachers agreed. They realized I was doing my job. The role of the principal required that I know what's going on in my school. Once teachers accepted the fact that I would visit their classrooms, my presence became no big deal.

And that's as it should be, because times are changing. As older principals retire, teachers will see a new breed of administrator who might even visit classes more often than I do. Supervision must be constant and aggressive. To know what's going on, the principal must be everywhere.

Evaluation, on the other hand

Although my experience with supervision turned out to be satisfactory, I'm less happy with my experience with evaluation. I've found the procedure used in my school system to be simplistic and unsatisfactory for fairly evaluating a teacher's performance. Indeed, the evaluation processes used in many school systems seem to have the same problems. And as a result, teachers often see evaluation as a painful experience.

The procedure my school system uses centers on a checklist on which principals must give each teacher one of four ratings: Outstanding, Satisfactory, Needs improvement, or Unsatisfactory. Our labor contract requires that teachers be evaluated formally once a year (three times if we're evaluating a probationary teacher), meaning most teachers get only one chance to make themselves look good to the evaluator.

For most teachers, this one-time evaluation is difficult, so I try to make it as painless as possible. I ask teachers to let me know when they would like me to visit. At one time, I distributed a form before the observation so teachers could tell me what to expect in their classroom, but teacher association officials complained I wasn't using the negotiated evaluation format, so I had to stop. Now teachers have no opportunity to explain formally in advance what I'm going to see.

During an evaluation, I generally write down in script format what I see happening in the classroom and then write several pages of observations and comments. I attach this to the district-required form.

Teachers generally react to my feedback in one of three ways, depending on their level of expertise: New teachers seem to welcome my suggestions. Some veterans accept my suggestions with a positive attitude but generally act relieved the process is finished for the year. Other teachers react negatively or spend the next few days in the faculty room rehashing my comments and counting the number of satisfactory checks they received compared to the teacher next door.

Once I've satisfied my contractual obligations, I generally revert to my supervisory stance for the remainder of the school year. With my informal visits, I try to keep tabs on whether teachers are following up on the suggestions I've made.

In my opinion, though, this one-shot evaluation procedure is essentially useless for improving instruction. Without a formal, negotiated procedure for mentors and peer coaches to follow up on evaluation findings, the evaluation is put aside to gather dust. I see little opportunity for improvement. An annual, 30-minute visit simply pays lip service to the notion of effective evaluation. Evaluation—like supervision—must be an ongoing look at what's happening in the classroom. Only then can we increase student achievement.

Four steps to better evaluation

Supervision and evaluation of staff are among the most important jobs a principal must accomplish, yet we tend to perform both perfunctorily. I think we need to change, and here are four steps to set us on the road to improvement:

1. ***Throw out checklists during evaluations.*** They don't work. Checklists make no provisions for improvement by the teacher. They are simply report cards. A better approach is a narrative report of a teacher's performance that describes exactly what's right and wrong about the teacher's techniques and sets targets for improved performance for the coming year.

2. ***Do away with annual formal observations.*** At the best of times, formal observations are an opportunity for the teacher to perform for the principal. I'm not certain an observation for which the teacher has had a chance to prepare is any real measure of what happens in that teacher's classroom day after day. Repeated visits give the principal a better perspective on the teacher's performance.

3. ***Provide opportunities for improvement.*** Simply telling a teacher what's wrong isn't enough. You need to provide opportunities for peer coaching that require teachers to observe one another in the classroom, critique what they see, and make suggestions that—because they don't come from a supervisor—don't seem threatening. Teachers need to attend workshops and return to the university classroom. Improvement comes from practice, not preaching.

4. ***Support these procedures.*** It's not enough to say you want change. You must work for it. You must push central office administrators to fund training for both principals and teachers, and you must lobby teachers to try peer coaching. You must set targets for your teachers' improvement, and you must get into the classrooms to see these targets are met.

Understanding the difference between supervision and evaluation—and knowing how to use both effectively—is the key to being a good principal. As principals, we don't always do the best possible job of either. But we can do better. D

SECTION 5

ADDITIONAL RESOURCES

ADDITIONAL RESOURCES

Evaluation Management Handbook by Magdala Raup and Francis Kolb, (1990), The Network, Inc., 300 Brickstone Square, Suite 900, Andover, MA 01810, (508) 470-0098.

Expanding Student Assessment by Vito Perrone, (1991), Association for Supervision and Curriculum Development, 225 N. Washington Street, Alexandria, VA 22314, (703) 549-9110.

Performance Assessment: What's Out There and How Useful Is It Really? by Judy Arter, a paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 1991), available from the Office of Educational Research and Improvement (OERI), Washington, DC 20202 or through ERIC, (212) 678-3433.

Portfolio Assessment in the Reading-Writing Classroom by Robert Tierney, et al, (1991), Christopher-Gordan Publishers, Inc., 480 Washington Street, Norwood, MA 02062, (617) 762-5577.